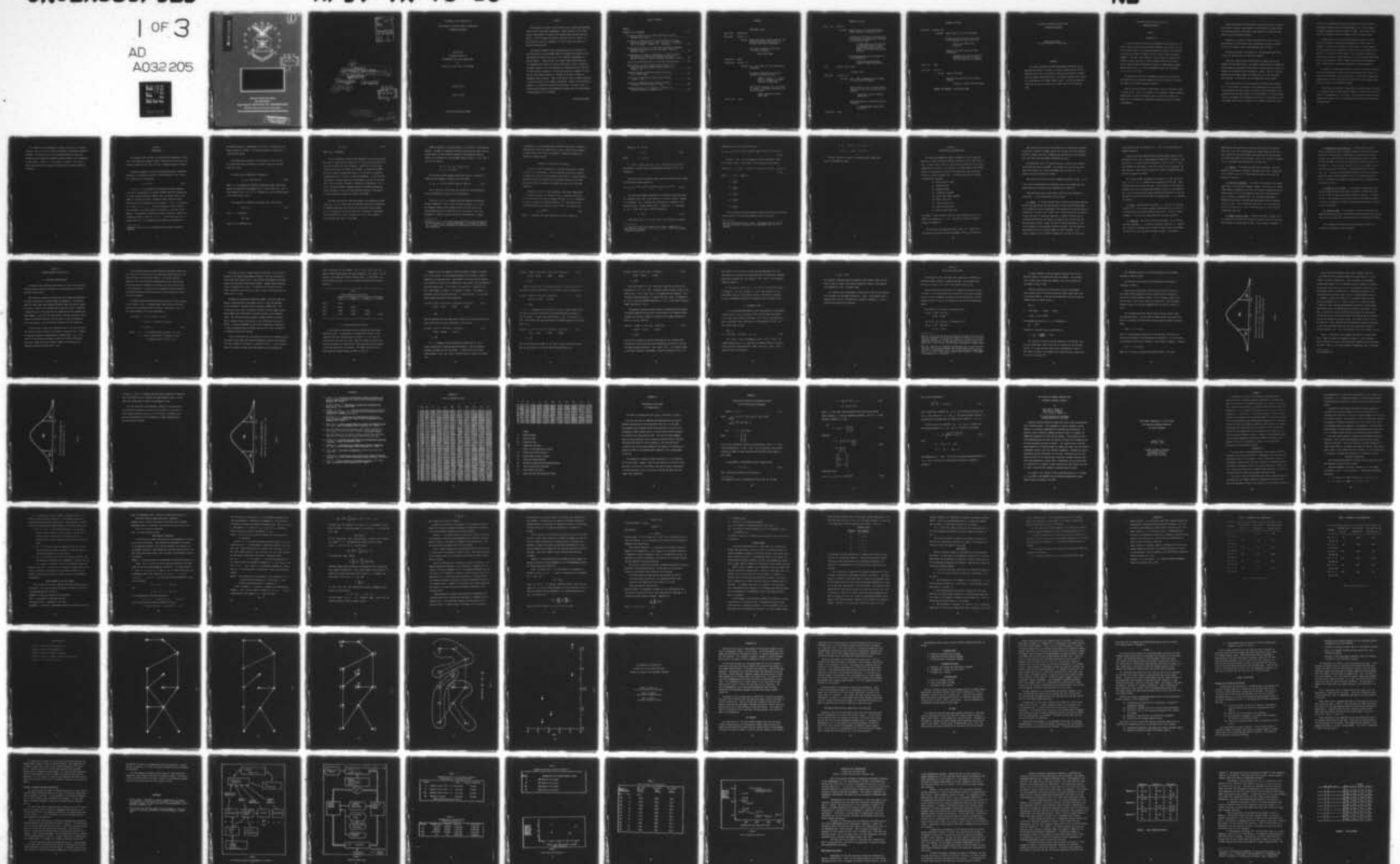
AD-A032 205 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OHIO F/G 12/2
PROCEEDINGS OF THE SYMPOSIUM ON APPLICATION OF DECISION THEORY --ETC(U)
OCT 76 N K WOMER

UNCLASSIFIED AFIT-TR-76-15 NL

1 OF 3
AD
A032 205



AD A032205

1

FG.

AF INSTITUTE OF TECHNOLOGY

UNITED STATES AIR FORCE

AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

6

PROCEEDINGS OF THE SYMPOSIUM ON APPLICATION OF DECISION THEORY TO PROBLEMS OF DIAGNOSIS AND REPAIR, Held at Fairborn, Ohio on 2-3 June 1976

14 AFIT-TR-76-15

10 Norman Keith Womer
Editor

11 October 1976

12 261p.

1473
012 200
LB

Proceedings of The Symposium on

APPLICATIONS OF DECISION THEORY TO PROBLEMS OF DIAGNOSIS AND REPAIR

Sponsored by

The Dayton Chapter of

THE AMERICAN STATISTICAL ASSOCIATION

and

The Air Force Institute of Technology

Fairborn, Ohio

June 2-3, 1976

Edited by Norman Keith Womer

Preface

The diagnosis and repair of malfunctions are of continuing importance in the fields of equipment maintenance. Recent advances in this field allow a large number of factors to be measured more precisely than ever before. If the increased information resulting from this trend is to yeild effective actions, advances of a similar nature are required in decision making as well.

The papers presented at this symposium focus on the analysis of decision making in an environment where the diagnosis and repair of complex mechanisms requires the use of still imperfect measurements of several factors. These analyses take several forms, among these are: (1) estimating the effects of incorrect decisions on costs, reliability and other elements of the loss function, (2) determining the variability of tests and test procedures, (3) determining the best sequence of tests, and (4) simulating the effects of alternative decision criteria on equipment repair processes. None of the papers offers a complete approach to decision making in this complex environment. However, each presents an approach to a significant portion of the problem. Our hope is that in discussing and comparing the approaches presented here, our understanding of the problem will be increased.

Norman Keith Womer

TABLE OF CONTENTS

PROGRAM

Wednesday, June 2

8:00- 8:30 Registration

8:30-10:00 Session I A

Welcome - Norman Keith Womer, Dayton Chapter of the American Statistical Association and Air Force Institute of Technology

"The Effect of WIPICS on the F4-B to N Conversion Program"

Norman Keith Womer

10:99-10:30 Break

10:30-12:00 Session I B

Chairman G. C. Saul Young, Air Force Institute of Technology

"The Effect of Renewal Processes Upon Stochastic Reliability Models"

Joseph E. Boyett, L. A. Dugas, and D. H. Hartmann, Air Force Institute of Technology

"Some Generic Properties of a Logic Model for Analysis of Hardware Maintenance and Design Concepts"

James P. Wong and William L. Andre, NASA

12:00- 1:30 Lunch

PROGRAM--Continued

1:30 - 3:30 Session II

Chairman Dwight Collins, Life Cycle Cost Working Group, Aeronautical Systems Division

"A Methodology for Analysis of Alternatives in the Acquisition and Support of Automatic Test Equipment Software"

E. James Dunne-Air Force Institute of Technology, Robert W. Morton-Aeronautical Systems Division, James L. Wilson-Electronic Systems Division

"An Experimental Design for Assessment of Test Repeatability"

Girard Levy and Horace Ray, Battelle Columbus Laboratories

3:30 NO HOST Coctail Party

Thursday, June 3

8:30-10:00 Session III A

Chairman Jon R. Hobbs, Management Science Center, Air Force Logistics Command

"Failure Detection Aids for Human Operator Decisions in Precision Inertial Navigation Systems"

Thomas Kerr, Analysis Sciences Corporation

"Analytical Models of Maintenance Decision Processes"

W. Stephen Demmy, Wright State University

10:00-10:30 Break

10:30-12:00 Session III B

Chairman Robert Gough, U.S. Air Force Academy

"Statistical Classification Methods with Special Emphasis on Growth Curve Data"

Jack C. Lee, Wright State University

"Diagnosis of Reliability Repair Stage and Remedies"

Edward Bilikam, Air Force Avionics Laboratory and Albert H. Moore, Air Force Institute of Technology

12:00- 1:30 Lunch

1:30- 3:30 Session IV

Chairman Norman Keith Womer

"Data Base Requirements for Distribution Studies"

Andrew Lai, Wright State University

SUMMARY AND COMMENTS - Norman Keith Womer

THE EFFECT OF WIPICS ON THE F-4B TO

N CONVERSION PROGRAM

Norman Keith Womer
Air Force Institute of Technology

ABSTRACT

This report provides the underlying theory and methods used to determine the effect of the Work in Process Inventory Control System (WIPICS) on the F-4B to N conversion program and the Naval Air Rework Facility, North Island, California. This report documents cost savings of 3.24 percent of the "before" WIPICS costs. It also concludes that these cost savings are statistically significant at the 10 percent level.

THE EFFECT OF WIPICS ON THE F4-B TO N CONVERSION PROGRAM

Section 1

INTRODUCTION

The report provides statistical documentation of the affect of the Work in Process Inventory Control System (WIPICS) on the Navy's F4-B to N conversion program. This program accomplishes a major life extension of the F4 at the Naval Air Rework Facility, North Island, California (NARFNI). The report draws heavily on previous work in evaluating the affect of complex systems on diverse programs especially the body of work on "Auditing Cost Effectiveness Analyses of Technological Changes" accomplished at the U. S. Naval Postgraduate School (see references [1, 3, 4, 5, 9, 11, 12]).

The purpose of this brief introductory section is to explain the philosophy of using estimated production and cost functions to document the benefits of a complex system like WIPICS.

Usually the major source of effectiveness for an information system is factor saving. That is, an analysis will frequently document expected effectiveness in terms of man-days saved or decreases in wasted raw materials or replacement of several more expensive pieces of equipment or procedures.

Other contributions to effectiveness are more difficult to estimate and evaluate. However, some attempt is usually made to include effects like increased speed of production, higher quality of output, and more control over the production process.

Frequently, an attempt is made to assign dollar values to each of these measures of effectiveness and a cost-effective change is defined as one with a greater value of effectiveness than its costs.

Frequently unstated in the analysis is the assumption that these effects are expected only if nothing else changes.

After the system has been installed one can gather data on the changes that have taken place. For example, we know what has happened to factor usage, we can measure the new speed of the production process, and perhaps we can determine quality changes in output. Here the important questions are not what changes have taken place, but why the changes have occurred. In particular are the changes due to the system?

Notice that after the fact, many other things may have changed as well. In particular, the outputs of the organization may have changed in both quantity and type, prices of factors, many have changed or new constraints may have been placed on the organization or old ones relaxed. Finally, the manager of the enterprise may not have acted exactly as the cost-effectiveness analysis expected. For example, suppose the system

replaces old equipment and several men, instead of laying off the men the manager reassigns them as trouble shooters. As a result, output increases dramatically with no increase in costs. Here costs did not decline as expected, instead the manager chose to increase output.

Current methods of auditing cost-effectiveness analyses involve looking at factors by categories and determining if their usage has changed. A portion of the change in factor usage is directly related to the system by verbal argument and a dollar value is assigned to that portion of the change in factor usage. Thus the relation between the factor savings and the system frequently is a verbal argument constructed by an outside observer. These "allowed" effects are then compared to costs. Other effects like quality changes are handled separately.

Our method of evaluating the affect of systems like WIPICS recognizes that an outside observer cannot possibly trace second order effects of many decisions through a massive enterprise. Thus we look at aggregated summary measures of the organization's behavior in each of several areas before and after the technological change.

In particular our approach to determining the effectiveness of WIPICS is to develop a model of the production behavior of NARFNI both before and after the implementation of the system. These production models are then compared for several sets of circumstances and conclusions are drawn from the comparisons.

This method has the advantage of allowing the analyst to ask many potential "what if" questions about the external circumstances affecting the NARF. It permits the use of well defined statistical tests and allows the use of generally recognized economic theory in the formulation of those tests. Finally, it is not nearly so subject to the spurious verbal description which frequently characterize some other methods of evaluation.

Section 2

METHODOLOGY

The purpose of this section is to describe the assumptions, limitations, and underlying production theory involved with determining a cost function to describe the F4-B to F4-N conversion program at NARFNI.

The basic assumption is that this production process is adequately described by a production function of the Cobb-Douglas[1] form. That is, for a given set of circumstances:

$$X = A\, L^{\alpha}\, M^{\beta}\, D^{\gamma} \tag{2.1}$$

In (2.1) X is a measure of the quantity of output produced. L, M, and D are measures of resources consumed identified respectively as direct man-hours expended, material cost in constant dollars, and number of aircraft days used. Production theory requires certain restrictions on the parameters to be estimated in (2.1). In particular $A > 0$ and $0 \leq \alpha, \beta, \gamma \leq 1$. The above production function implies that increases in any input permits an increase in the output of the process. The production function also permits decreasing, constant, or increasing returns to scale as $\alpha + \beta + \gamma$ is less than, equal to, or greater than one. This formulation also implies that production in

[1]Sometimes referred to as the "Generalized Cobb-Douglas Production Function."

the Beeline program is independent of the level of production in the other programs at the NARF. The latter assumption is relaxed to a limited extent below.

The second major assumption in the analysis is that the NARF has control only over the quantity of inputs it uses and not over their unit costs.

The total cost of production is defined as:

$$TC = C_o + DL\$ + M\$ + P\$ \qquad (2.2)$$

where C_o is overhead cost including production expense applied and general and administrative expense, DL\$ is direct labor cost, M\$ is direct material cost, and P\$ is a penalty cost for the imputed value of an aircraft *day in shop.*

The assumption of exogenous unit factor costs implies that:

$$DL\$ = P_\ell \cdot L \qquad (2.3)$$

where P_ℓ is exogenous;

$$M\$ = P_m \cdot M \qquad (2.4)$$

where P_m is exogenous; and

$$P\$ = P_d \cdot D \tag{2.5}$$

where P_d is exogenous.

Like all assumptions these are not expected to be fulfilled exactly but there is some reason to expect that to a large extent the NARF does not exert control over these variables. Clearly Civil Service wage scales are not under the control of the NARF. In addition the NARF's ability to adjust the experience level of its personnel to alter the average wage rate is, at best, a long run phenomenon. In any case this fluctuation in average wage rate is more usually a result of Navy wide policy than of internal response to production circumstances. This is not to say that the NARF is totally unable to influence average wage rates, but merely to point out that the major sources of fluctuation in this variable are outside the control of the NARF.

The other two variables that are assumed to be exogenous are more clear cut. P_m is a price index for material costs that is determined by economic conditions and the policies of the Naval Supply system. P_d is an imputed value of an aircraft day in shop and is determined by the usefulness of a day's service by an F4-N. This variable is also not under the control of the NARF.

Another assumption is used to derive a cost function for the Beeline program. The NARF is presumed to minimize the cost of any given level of output subject to the constraint imposed by the production function. Formally the assumption is that the NARF chooses values of L, M. and D to solve the program:

$$\begin{aligned} \text{Min} \quad & TC = C_0 + P_\ell \cdot L + P_m \cdot M + P_d \cdot D \\ \text{s.t.} \quad & X = A\, L^\alpha M^\beta D^\gamma \end{aligned} \tag{2.6}$$

The solution to this program yields total cost as a function of output and the exogenous variables as defined below.

$$TC = C_0 + A' \, X^{1/\alpha+\beta+\gamma} \, P_\ell^{\alpha/\alpha+\beta+\gamma} \, P_m^{\beta/\alpha+\beta+\gamma} \, P_d^{\gamma/\alpha+\beta+\gamma} \tag{2.7}$$

This is the basic form of the cost function for the Beeline program corresponding to a given set of circumstances.

The use of (2.7) as a reduced form cost function involves one further assumption.[1] That is that the level of output, X, is also an exogenous variable to the system, specified by the fleet or at least determined prior to beginning work on the project. This last assumption is subject to some criticism for some measures of output.

---

[1] See Nerlove [8] for a more formal elaboration of these assumptions and the role they play in model construction. The comments by Dhrymes [2 , p. 234] on the requirement for imperfect cost minimization are also appropriate.

In particular in an environment where production load norm is renegotiated during or after the completion of a job; it would not be a measure of output which satisfied this assumption. Alternative measures of output are suggested below.

Changing Circumstances of Production

Up to this point the derivation of the cost function has presumed a given set of circumstances. Two circumstances which may be expected to change over time are the efficiency of labor due to a learning effect and the efficiency of production on the Beeline program due to interaction with other F4 programs at the NARF which use some of the same resources.

To handle the first of these changing circumstances suppose that L, the labor referred to in equation (2.1), represents labor that is as efficient as that used to produce the first B to N conversion. If labor is subject to a learning effect of "a" then labor in the first conversion may be related to labor in the n th conversion by

$$L_n = N^{a/\ell n 2} L \qquad (2.8)$$

where L represents the labor required in the first conversion.

Rewriting (8) we have

$$L = L_n N^{\delta} \tag{2.9}$$

where[1] $\delta = - a/\ell n\ 2$

If (2.9) is substituted into (2.1) the production function is now defined in terms of labor of appropriate efficiency for the n th conversion.

Following through the analysis above the cost function may be changed to reflect the learning effect as:

$$TC = C_o + A' X^{1/\alpha+\beta+\gamma} P_{\ell}^{\alpha/\alpha+\beta+\gamma} P_m^{\beta/\alpha+\beta+\gamma} P_d^{\gamma/\alpha+\beta+\gamma} N^{-\alpha\delta/\alpha+\beta+\gamma} \tag{2.10}$$

The interaction between production on the Beeline program and other F4 programs at the NARF can be handled in a similar fashion. Suppose A in equation (2.1) represents the efficiency of production in the Beeline program. Further suppose, that A depends on K, the number of F4's in the 953 shop, in the following way.

$$A = B K^{\omega} \tag{2.11}$$

Substituting from (2.11) into (2.1) and following the analysis

---

[1] a, the natural log of the learning curve slope is expected to be less than 0, therefore $\delta > 0$, i.e. $\ell n\ .8 = - 0.223$ for an eighty percent learning curve.

above the cost function may be written as:

$$TC = C_o + B' X^{1/\alpha+\beta+\gamma} P_\ell^{\alpha/\alpha+\beta+\gamma} P_m^{\beta/\alpha+\beta+\gamma} P_d^{\gamma/\alpha+\beta+\gamma} N^{-\alpha\delta/\alpha+\beta+\gamma} K^{-\omega/\alpha+\beta+\gamma} \tag{2.12}$$

Equation (2.12) may be estimated in either the before WIPICS or the after WIPICS environments by the following transformation:

$$\log(TC-C_o) = \beta_o + \beta_1 \log X + \beta_2 \log P_\ell + \beta_3 \log P_m + \beta_4 \log N + \beta_5 \log K + \varepsilon \tag{2.13}$$

where[1] $\beta_o = \log B' + \frac{\gamma}{\alpha+\beta+\gamma} P_d$

$$\beta_1 = \frac{1}{\alpha+\beta+\gamma}$$

$$\beta_2 = \frac{\alpha}{\alpha+\beta+\gamma}$$

$$\beta_3 = \frac{\beta}{\alpha+\beta+\gamma}$$

$$\beta_4 = \frac{-\alpha\delta}{\alpha+\beta+\gamma}$$

$$\beta_5 = \frac{-\omega}{\alpha+\beta+\gamma}$$

To be consistent with the assumptions above the following restrictions on the values of the estimated parameters must be satisfied:

[1]Since all observations are on F4-N's, the penalty cost per aircraft day does not change from observation to observation and P_d is a constant.

$$0 < \beta_0 \quad , \quad 1/3 \leq \beta_1 \quad , \quad 0 \leq \beta_2 \leq 1 \quad ,$$

$$0 \leq \beta_3 \leq 1 \quad , \text{ and } \quad -1 \leq \beta_4 \leq 0.$$

The last restriction allows for learning curves ranging from fifty to one-hundred percent.

Section 3

DESCRIPTION OF AVAILABLE DATA

The preceding mathematical model is defined in terms of physical measures of inputs and outputs and well defined prices for these quantities. In fact, data on these variables cannot be obtained in most large organizations and NARFNI is no exception. The data obtained from NARFNI consists of information on 83 F4-B to N conversions begun during the period July 1972 to December 1973. For each conversion the following information was available:

(1) Identification number

(2) Induction data

(3) Production date

(4) Production load norm

(5) Direct labor hours expended

(6) Direct labor cost

(7) Direct material cost

(8) Applied overhead cost

(9) Applied statistical cost

The manner in which the above statistics are accumulated and used by NARFNI are enumerated in [6]. The data is listed for reference purposes in Appendix A.

The induction and production dates of each F4 rework job in shop during the period July 1972 to December 1973 was also provided.

Also available was data on the standard hours produced and the number of man hours required to produce them by the 953-shop (aircraft assembly). Prior to 1 January 1973 this information was available only on a quarterly basis but after that time weekly information was used.

One additional item of information concerns a penalty cost assigned to aircraft down time at the NARF. This penalty cost per day was calculated from average unit flyaway procurement cost for the F4-J. The penalty cost was calculated as follows:

$$[\text{Unit cost of F4-J/Service Life}] = [\text{Penalty cost/day in shop}] \quad (3.1)$$

A unit cost of $2,492,000 and an expected service life of 2400 days (80 months) were used to obtain a daily penalty cost of $1038.23.

These data were used to form several proxies for the variables of Section Two. A brief description of each follows:

1. <u>Output</u>. If it were the case that an F4-B to N conversion required exactly the same repairs and modifications for each aircraft there would be no problem with output measurement. The completion of each job requires bringing each aircraft up to a given level of modification and an extensive inspect and repair as necessary operation. Different aircraft require different modifications; and repairs are made on some aircraft that are not necessary on others. Hence, the level of effort expended on a conversion depends on the particular aircraft involved. Thus the output of the conversion effort is not the number of F4-N's produced. It is rather a measure of the difference between the F4-B that the fleet gave

up and the F4-N that was returned to it. This is the output which we attempt to measure.

Clearly none of the data available directly measure output of the conversion effort, hence a proxy measure of some kind is necessary. One possibility is to merely ignore differences in jobs and act as if each job produced the same output. While the results of this assumption are also reported, a proxy variable based on an index of standard hours was used as our measure of output.

This index attempts to measure work content of a given job by the class A, B, C, or D standard hours associated with the job. These standards tend to be stable over time. Thus increases in the number of standard hours tend to indicate more extensive repairs or modifications for a job and therefore more output produced upon completion of the job. *The method for calculating the index of standard hours,* S, *is reported* in Appendix B.

<u>2. Labor</u>. Direct man-hours expended, L, are used as a measure of labor used. This variable excludes indirect man-hours and fails to distinguish among alternative types of labor used. In spite of these drawbacks, it is one of the most accurate of the proxy variables.

<u>3. Wage rate</u>. The ratio of direct labor costs, L\$, to direct man-hours expended, L, is used as a proxy for wage rate P_ℓ. Because this variable is averaged over all types of labor it may be influenced by factors other than a general increase in wages. For example a

modification that requires very technical work and uses high priced labor exclusively will be associated with higher values of this proxy. Also, periods of time that involve large quantities of overtime labor will show high values of the variable. Still, this variable is thought to be a reasonably accurate measure of the wage rate.

4. Material. No attempt has been made to construct a physical measure of the quantity of material consumed on each job. Instead the cost of material used at FY 1973 prices was used as a proxy for the quantity of material.

5. The Price of Material. An index of the material price change experienced 1 July 1973 was formed as follows: The value of all identifiable material expended on the F4-B to N conversion program during the fourth quarter of FY 73 was compared to the cost of that material at prices in effect 31 December 1974. This difference \$453,752.92 - \$438,333.58 was used to calculate the percentage increase in material prices as 3.52%. Thus the material price index, P_m, our proxy for the price of material is 1.00 for periods prior to 1 July 1973 and 1.0352 after that time.

6. Number of days in shop. It seems clear that if output is to be a measure of modifications and repairs to F4-B's then one input must be the number of aircraft days in shop. This variable is denoted D.

7. Penalty cost per aircraft day. The NARF does not pay for F4-B's to convert nor does it provide "loaners" to the fleet while the conversion is taking place. Nevertheless, from the point of view of the Navy, the pipeline of F4-N's in the conversion process is costly and both the NARF and the Navy are concerned with reducing the size of the pipeline. Hence our study includes as costs of the NARF a penalty cost per aircraft day of $1038.23. As explained above this penalty cost, P_D, is based on the procurement cost of an F4-J; the assumption being that those aircraft replacing F4-B's in the fleet were F4-J's not B's. Therefore the appropriate opportunity cost is related to the procurement of F4-J's.

8. Number of F4's in shop. In the attempt to capture any interactions between the Beeline and other programs at the NARF the number of F4's in the 953 shop at any point in time was measured. The number of F4's in shop averaged over the days that any given Beeline job was in shop was included as a variable in the regression analysis to account for this interaction.

9. Sequence number. In an attempt to correct for learning that may have been taking place as the Beeline program progressed the sequence number of each job was also recorded and used in the regression analysis.

These aggregated data were used as described in Section 4 to estimate the parameters of the cost models.

Section 4

ESTIMATION AND HYPOTHESIS TESTS

A. Preliminary Data Analysis

The nature of the available data described in the previous section dictated certain regression techniques in the estimation of the before and after cost function.

While data was provided on 83 jobs only 52 of these were completed. The other 31 jobs were in various stages of completion. The data provided for these later jobs contained estimated costs and time to completion and as a result it was decided not to make use of this information. In addition one of the 52 jobs that was completed had been inducted with crash damage. It was felt that the nature of the work involved on that job was sufficiently different from the others to justify dropping it from the data base. Thus 51 jobs were available for the estimation.

Of the 51 jobs 25 were still inprocess after 1 July 1973, the date the chaining feature of WIPICS first became available for use. However, no data was available on completed jobs inducted after 1 July. As a result it was decided to estimate a single cost function for Beeline jobs with a dummy variable for WIPICS instead of two separate cost functions one before and one after 1 July.

This procedure requires a specification of how WIPICS affects the cost function (in this case by a one time shift) and therefore is less able to account for more subtle affects. On the other hand this approach avoids the allocation of observations to two functions and the autocorrelation introduced by prorating jobs over time required by the alternative. It also avoids the hypothesis testing problem in environments where the sample variance may differ between the before and after situation.

The dummy variable was defined as the proportion of time a job was in shop before 1 July. A second variable (one minus the first one) expresses the proportion of time a job was in shop after 1 July. By this device equation (2.13) was transformed to:

$$\begin{aligned} \log (TC-C_0) &= \beta_b W + \beta_a (1-W) + \beta_1 \log X \\ &\quad + \beta_2 \log P_\ell + \beta_3 \log P_m + \beta_4 \log N \\ &\quad + \beta_5 \log K + \varepsilon \qquad (4.1) \end{aligned}$$

where $W = 1$ if production date is before 1 July and

$W =$ (days in shop before 1 July)/days in shop if production date is after 1 July.

This does not yield a linear effect of time after 1 July on costs because of the convex transformation involved. This has the effect of giving jobs only slightly after 1 July slightly more emphasis than they deserve thus we penalize the system slightly. However these considerations are expected to be minor. The approach also assumes that the effect of WIPICS on cost savings of jobs is evenly distributed over the time a job is in shop.

The data also created one additional change. The price index for material experienced only one change, on July 1, 1973 the estimated prices of material increased 3.52% as computed by NARFNI [7]. Thus the material price index is perfectly collinear with the dummy variable due to WIPICS and coefficients for neither are able to be estimated alone. Thus the cost function contains a dummy variable for effects due to discreet changes occurring on 1 July 1973 and not just due to WIPICS. A method recommended by Theil [10] as explained in Appendix C is used to separate these two affects and the price index for material is dropped from the cost function.

Prior to estimating the cost function the matrix of estimated correlation coefficients among the proposed independent variables was examined. The lower triangle of the matrix is recorded in Table 1. From an examination of the table it seems clear that several of the variables are

highly correlated with one another. That is log X, log P_ℓ, and log K seem to contain approximately the same information. This result is quite obvious in the regression results reported below. It also seems clear that the dummy variable, W, is only slightly correlated with the other variables.

TABLE 1

SIMPLE CORRELATION COEFFICIENTS

	W	log P_ℓ	log N	log K
log P_ℓ	-0.175			
log N	-0.691	0.497		
log K	-0.055	0.991	0.402	
log X	-0.116	0.987	0.385	0.991

B. Estimating the Cost Function

In the process of estimating the cost function one fact became painfully obvious. No matter which variables were included in the regression the first observation was an outlier. In all cases its residual error was more than three times the standard deviation indicating an extremely low probability of that observation being a member of the same population as the others. For these reasons the first observation was dropped leaving a sample of size 50.

Because of the high degree of multicollinearity attempts to include all of the variables in the regression proved to be fruitless. Several regressions are reported below. Each contains a different set of explanatory variables for the log of variable cost, direct labor cost plus material cost plus penalty cost. In all cases, the regression is reported first with, then without including X our measure of output. The numbers in parenthesis below the coefficients are t statistics and S is the estimated standard deviation of the regression.

$$\log(VC) = \underset{(5.29)}{2.94}\, W + \underset{(5.18)}{2.92}\,(1-W) + \underset{(0.50)}{0.20}\,\log P_\ell + \underset{(9.07)}{0.61}\,\log X \qquad (4.2)$$

$S = .022$

Here the coefficient of the wage rate is clearly not significant while the other coefficients are each significant at the 5% level.

$$\log(VC) = \underset{(14.81)}{7.21}\, W + \underset{(14.46)}{7.24}\,(1-W) - \underset{(-3.01)}{1.75}\,\log P_\ell \qquad (4.3)$$

$S = .036$

If X is dropped from the relation the coefficient of $\log P_\ell$ changes dramatically indicating multicollinearity. Also the standard deviation increases by over 50 percent. It seems clear that we must choose between $\log X$ and $\log P_\ell$ and the choice is clearly to include $\log X$.

$$\log(VC) = \underset{(9.14)}{3.00}\, W + \underset{(8.76)}{2.97}\,(1-W) + \underset{(0.86)}{0.01}\,\log N + \underset{(8.64)}{0.61}\,\log X \qquad (4.4)$$

$S = .022$

Here the coefficient of the sequence number is clearly not significant while the other coefficients are significant at the 5 percent level.

$$\log(VC) = \underset{(235.92)}{5.83}\, W + \underset{(127.85)}{5.89}\,(1-W) - \underset{(-3.65)}{0.078}\,\log N \qquad (4.5)$$

$S = .035$

If X is dropped from the relation we do get a negative coefficient for $\log N$ as we would expect but the standard deviation increases by over 50 percent. It seems clear that relation (4.4) is preferable to (4.5) and likewise that the contribution of sequence number to explaining cost is nil.

$$\log(VC) = \underset{(2.01)}{1.71}\, W + \underset{(2.08)}{1.73}\,(1-W) + \underset{(1.82)}{0.66}\,\log K + \underset{(9.69)}{0.67}\,\log X \qquad (4.6)$$

$S = .021$

Here the coefficient of number of F4 jobs in shop is not quite significant at the 5 percent level while the other coefficients are.

$$\log (VC) = 8.23\ W + 8.14\ (1-W) - 1.42 \log K \qquad (4.7)$$

$$(9.24) \quad (9.43) \quad (-2.80)$$

$$S = .036$$

While the results of (4.6) are not by themselves disturbing the comparison of (4.7) to (4.6) indicates a rather dramatic change in the coefficient of log K when log X is removed from the equation. Thus indicating multicollinearity. It seems clear that log X is preferable to log K since the standard deviation of (4.7) is more than 50 percent higher than that of (4.6).

The previous regression results lead to a formulation which because of multicollinearity and the lack of significance of the sequence number includes only the log of output in addition to the dummy variables for WIPICS in the cost function.

$$\log (VC) = 3.186\ W + 3.174\ (1-W) + 0.595 \log X \qquad (4.8)$$

$$(12.99) \quad (12.92) \quad (10.42)$$

$$S = .0219$$

Here all the variables are easily significant at the 5 percent level and despite the fact that only three explanatory variables are used the estimated standard deviation is low. Hence (4.8) is at least as good as the other relations investigated in terms of goodness of fit, it is

the simplest of the relations and the sign and magnitude of all the coefficients are believable and consistent with the theoretical considerations which underlie the cost function. Thus (4.8) is the preferred regression equation.

R^2 for the last regression is .70 and R^2 adjusted for degrees of freedom is .68. The residuals of (4.8) were examined for autocorrelation using the Durbin-Watson statistic and the null hypothesis of serial independence failed to be rejected at the 95 percent level.

C. Hypothesis Tests

It is a straightforward matter to test the hypothesis that expected variable costs in the period after 1 July 1973 are lower than before 1 July other things remaining the same. This is just a one tailed test that a particular linear combination of the parameters of (4.8) are less than or equal to zero.

That is:

$$a = E\left(\log (VC_b) - \log (VC_a)\right) = \beta_b - \beta_a \qquad (4.9)$$

From (4.8) a may be estimated as 3.186 - 3.174 = 0.0124. The standard deviation of $\hat{a}$, S_a may also be estimated from the variance-covariance matrix associated with (4.8). Here S_a = .0110. Thus a t statistic may be calculated as

$$t = \hat{a}/S_a = 1.125$$

This statistic indicated that the hypothesis that "before" costs are less than or equal to "after" costs must be rejected in favor of the alternative hypothesis at the 15 percent level.

This result can be strengthened by removing the affect of the material price increase from the dummy variables in (4.8). The procedure used is based on Theil [10, pp. 147-155] and is reported in Appendix C. These results from the basis for Section 5.

Section 5

RESULTS AND CONCLUSIONS

The probability that there were cost savings due[1] to WIPICS significantly greater than \$0 is greater than 90%. The estimated cost savings due to WIPICS are 3.24% of the "before" WIPICS costs.

These conclusions are reached by the regression analysis reported in Section 4 which determines a cost function for the Beeline program in each of two environments; one before the chaining feature of WIPICS became available for use on the program, and one after the feature was usable.

The "before" cost function is estimated to be:

$$\log C_B = \underset{(12.99)}{3.186} + \underset{(10.42)}{.595} \log X$$

The "after" cost function is estimated to be:[2]

$$\log C_A = \underset{(12.92)}{3.174} + \underset{(10.42)}{.595} \log X$$

---

[1]That is, the cost savings were associated with an event which coincided with the introduction of the chaining feature of WIPICS to the F4-B to N conversion program. Furthermore this effect was in addition to effects associated with changes in workload, wage rates, material prices, the number of F4 jobs in shop, and learning as expressed by sequence number.

[2]The "two" functions are estimated simultaneously with the specification that they differ from one another only by the constant term. The coefficient of determination, R^2, is .70 and the estimated standard deviation, S, is .0219. These results apply to both functions. The numbers in parentheses are t statistics.

In these relations C_A and C_B represent variable costs of the Beeline program in the before and after environments. The variable costs include direct labor cost, material cost, and a penalty cost for number of days in shop.

From the estimated cost function one can test the hypothesis that variable costs in the after period are significantly lower than in the before period. To do this we estimate the log of the ratio of "before" costs to "after" costs, a.

That is:

$$a = \log (C_B/C_A) = \log C_B - \log C_A$$

$$= 3.186 - 3.174 = 0.0124$$

The standard deviation of a, S_a is estimated as:

$$S_a = .0110$$

Therefore a t statistic may be calculated as:

$$t = \frac{a}{S_a} = \frac{0.0124}{0.0110} = 1.125$$

This statistic indicates that the hypothesis that "before" costs are not greater than "after" costs must be rejected at the 15% level. That is, if it were true that there are no improvements in costs in the "after" situation, the probability of obtaining such a large value as 1.125 is less than 15%.

The estimated value of a at 0.0124 corresponds to an estimated decrease in costs of 2.82%.

The empirical distribution of the percentage cost decrease is indicated in figure 1.

These results indicating a significant cost savings of 2.82% in the "after" period understate the case concerning the influence of the WIPICS chaining features on costs. This is because a major influence on costs in the "after" period, the affect of a 3.52% in average material prices, has not been separated from the affect of WIPICS on costs.

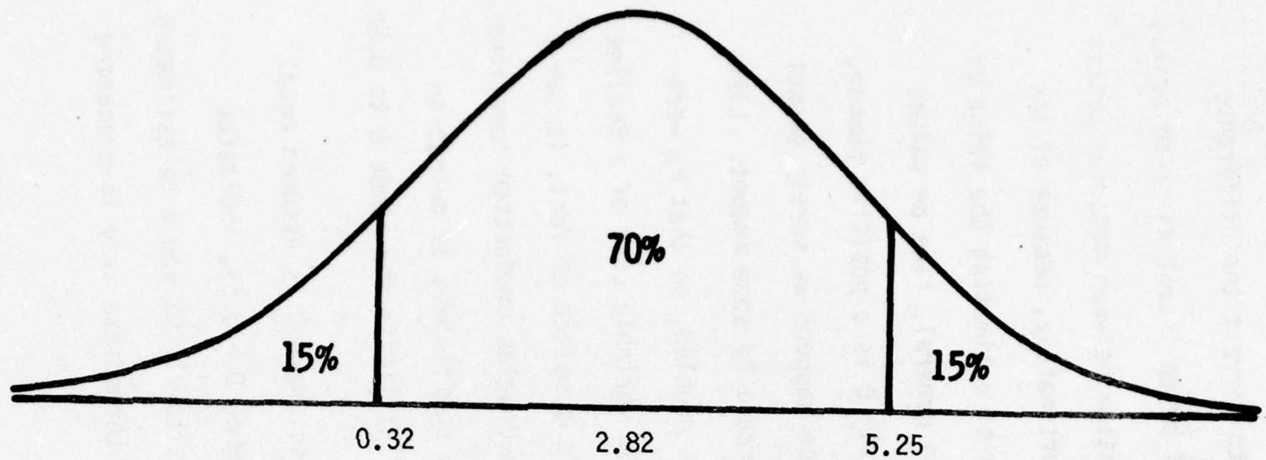
The following sensitivity analysis may be used to unravel these two entwined affects. The fact that the dummy variable associated with WIPICS *is perfectly collinear with the price index* for material requires that:

$$3.186 = B + D \log P_B$$

That is, the estimated constant term in the "before" cost function is not just an estimate of the productivity before WIPICS. It also includes the influence of the price of material in the "before" situation. Likewise,

$$3.174 = A + D \log P_A$$

Where P_A is an index of the price of material after 1 July 1973.



Distribution of estimated percentage cost decrease per job since 1 July 1973. (The probability that there were no cost decreases is less than 15%).

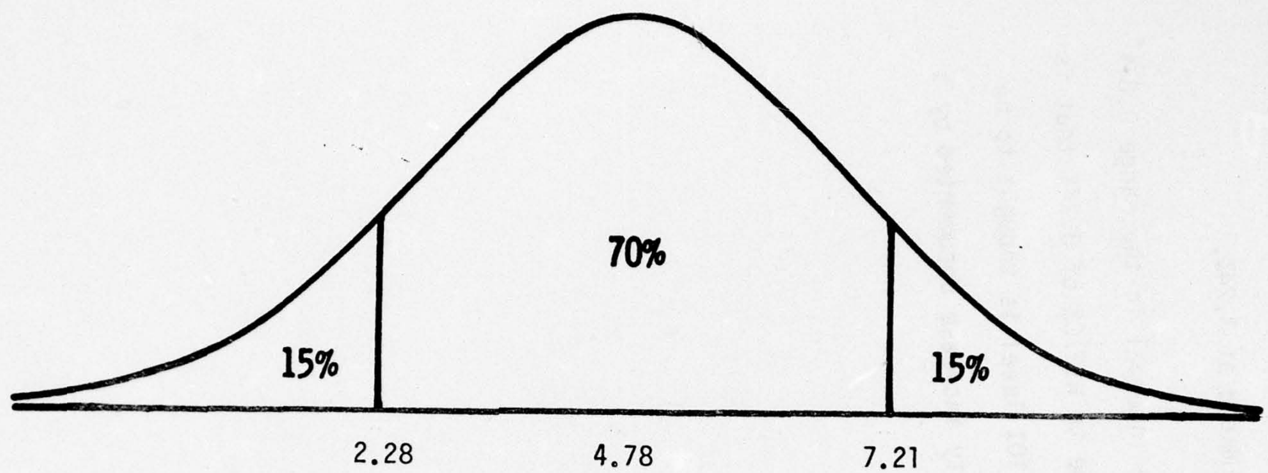
Figure 1

Since we know that material prices have increased 3.52% since 1 July 1973[1] we can with perfect generality assign values of 1.0 and 1.0352 to P_A and P_B respectively. Thus, to correct the difference between these numbers for the influence of price increases on material, we need only know the value of D, the relation between material prices and variable costs of a Beeline job. Unfortunately, because of the multicollinearity problem, there is no way of estimating the value of D. However, it seems clear that D will, in general, take on values only in the range of 0 to 1. That is clearly D is a positive number, if material prices increase and nothing else happens we surely expect the variable costs of a Beeline job to increase by some amount. Likewise, it seems clear that if material costs doubled, so that P_A were one and P_B were two, we would not expect the variable cost of a Beeline job to also increase by a factor of two. As a matter of fact, it can be shown that if the assumption of the Cobb-Douglas production functions is met, then the maximum value for D is the coefficient of output in the cost function, .59. Thus D is expected to lie in the range 0 to 0.59.

Figure 1 corresponds to the extreme case where D is assumed equal to 0. Figure 2, based on the opposite extreme, D = 0.59, indicates that if this were the case, cost savings due to WIPICS would be estimated at 4.78%. Finally, a far more believable intermediate case is presented

---

[1]See reference [7].

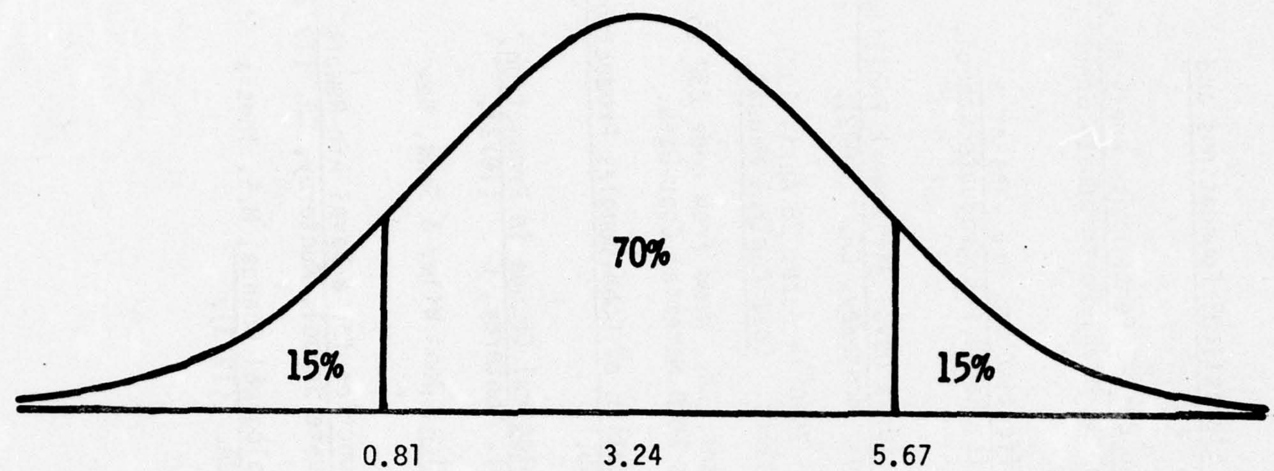


Distribution of estimated percentage cost
decrease due to WIPICS with D = .59.

Figure 2

in Figure 3. Here D is assumed to be equal to the proportion of material cost to variable cost for a Beeline job, approximately 0.125. In this case, cost savings due to WIPICS are estimated at 3.24%.

This last case with a 70% confidence interval in the range 0.81% to 5.67% and an estimated cost savings due to WIPICS of 3.24% that is significantly different from zero at the 10% level is thought to be the best representation of the productivity increase documented by a cost savings due to WIPICS.



Distribution of estimated percentage cost decrease per job due to WIPICS with D = .125.

Figure 3

REFERENCES

1. Bradley, C.W., Estimation of Cobb-Douglas Production Function at a Naval Air Rework Facility, M.S. Thesis, Naval Postgraduate School, Monterey, CA. (1972).

2. Dhrymes, Phoebus J., Econometrics: Statistical Foundations and Applications, Harper & Row, 1970.

3. Hartman, J.K. and _______. Auditing Cost-Effectiveness Analyses of Technological Changes, Technical Report, Naval Postgraduate School, Monterey, CA. NPS55HH73121A (1973).

4. McGarrahan, J.R., Comparative Cost Effectiveness Analyses at a Naval Air Rework Facility, M.S. Thesis, Naval Postgraduate School, Monterey, CA. (1973).

5. Myers, W.M., A Linear Economic Model for a Naval Air Rework Facility, M.S. Thesis, Naval Postgraduate School, Monterey, CA. (1972).

6. Naval Air Rework Facility Instruction 7650.1A dates 28 April 1971 with Change 1 dated 30 July 1971; Subject: Cost Control Manual.

7. Naval Air Rework Facility, North Island, CA., memo from code 232 to code 210 of 12 February 1974, Subject F4-N Material Cost Data.

8. Nerlove, M., Estimation and Identification of Cobb-Douglas Production Functions, Chicago, Rand McNally, 1965.

9. Spooner, R.L., Evaluation of a Technological Change in Production, M.S. Thesis, Naval Postgraduate School, Monterey, CA. (1972).

10. Theil, Henri, Principles of Econometrics, John Wiley & Sons, New York, 1971.

11. Trafton, W.C., Estimation of a Cost Function for a Naval Air Rework Facility, M.S. Thesis, Naval Postgraduate School, Monterey, CA. (1972).

12. Tye, D.L., Auditing Analyses of Technological Change, M.S. Thesis Naval Postgraduate School, Monterey, CA. (1973).

APPENDIX A

F4-B to N Conversion Data

K	ID	PD	N	L	L$	M$	PE	GA	NC	SC	TC
1	2183	2355	34346	34403	225876	59087	99674	141676	526313	220010	746323
2	2187	3081	29500	38112	251785	51750	110221	160242	573998	228428	802426
3	2196	3075	41672	41717	274861	58879	123167	176153	633060	226492	859552
4	2215	3081	36252	41472	275288	84975	125035	177124	662422	283548	945970
5	2223	3089	39177	39666	263031	60807	113793	173078	610709	227066	837775
6	2231	3117	40462	40462	271404	85495	114126	179411	650436	198382	848818
7	2238	3117	40783	40783	275167	73879	117683	182711	649440	230452	879892
8	2245	3134	40220	40220	271970	83419	114296	181513	651198	274465	925663
9	2255	3144	40130	40130	273723	95217	117579	181557	668076	278714	946790
10	2258	3149	29500	38354	278233	62368	113167	176849	630617	208864	839481
11	2264	3177	42103	42178	290183	72895	125876	195442	684396	266180	950576
12	2270	3151	31000	32026	218319	60272	93261	148714	520566	221580	742146
13	2277	3120	29100	31485	213561	59411	92808	143418	509198	110310	619508
14	2280	3128	34100	33586	229457	66889	99825	154874	551045	135837	686882
15	2286	3151	29100	36218	251945	57404	109211	168712	587272	151968	739240
16	2290	3162	28600	31339	214480	58157	94170	146730	513537	128416	641952
17	2292	3151	28800	36239	249338	59937	107794	170246	587315	137374	724689
18	2298	3151	27600	31203	217535	55954	92221	147406	513116	159433	672549
19	2300	3169	26600	28891	202092	53899	83855	137839	477685	108413	586098
20	2306	3212	40812	42267	290390	59523	123176	203286	676375	155714	832089
21	2311	3173	28100	30412	213362	54784	92372	146129	506647	226142	732789
22	2314	3166	27100	31463	223109	48810	89836	151116	512871	138247	651118
23	2319	3179	28100	33950	239320	54734	100184	164031	558269	208502	766771
24	2322	3166	29100	36074	255034	55997	103962	174361	589354	205340	794693
25	2326	3166	28100	31020	218843	52296	86042	150429	507610	214239	721849
26	2332	3179	29100	37874	266316	58966	103844	184866	613992	210340	824332
27	2335	3180	29100	34543	244793	57191	103121	169547	574652	220358	795010
28	2341	3234	34846	36066	250786	59628	103372	175817	589603	216512	806115
29	2346	3270	32100	32017	223837	70031	101469	153025	548502	229419	777921
30	2350	3257	32100	35697	249888	62750	103286	173830	589754	212059	801813
31	2355	3271	32100	34618	246988	69858	107459	167401	591706	256002	847708
32	3003	3262	32100	33451	235555	70632	100094	163059	569340	184430	753770
36	3025	3264	32100	28633	203392	67802	79920	138825	489939	84790	574729
37	3032	3269	35300	31422	224856	71440	86377	152957	535630	204941	740571
39	3044	3236	32100	23285	168276	69873	67234	113876	419259	126889	547148
40	3054	3271	38300	34506	247579	87221	101453	167300	603553	307551	911104
38	3038	3346	39136	36069	254080	92532	107306	171441	625359	354297	979656
43	3067	3271	38300	34253	245704	99742	104659	164837	614942	344061	959003
42	3064	3271	32100	30409	219235	74415	90447	145754	527851	207156	735007
44	3073	3271	32100	28431	205083	79415	86477	135991	506966	271398	778364
41	3059	3257	32100	27900	199042	76636	87533	135841	499052	173286	672338

K	ID	PD	N	L	L$	M$	PE	GA	NC	SC	TC
41	3059	3257	32100	27900	199042	76636	87533	135841	499052	173286	672338
45	3078	3277	35300	31854	227894	88338	95915	153884	566031	349739	915770
46	3081	3302	32100	28032	197032	85820	86226	133482	502560	184797	687357
47	3096	3299	38300	32216	223585	68251	103476	153152	548464	306557	855021
48	3102	3340	32100	30043	212786	55362	96089	142168	506405	91984	598389
49	3108	3351	32100	35687	253071	70387	112982	168136	604576	107939	712515
50	3117	3361	32100	36724	259648	82807	114909	172803	630167	128731	758898
51	3127	3341	38300	34107	246276	67423	106340	160110	580149	250583	830732
52	3136	3361	32100	34027	244312	69805	108173	158570	580860	93589	674449
53	3144	3362	35300	35181	251545	59262	113280	161476	585563	305146	890709
54	3156	3362	38300	36201	260407	64213	113226	166823	604669	261530	866199

	Legend
K	Sequence Number
ID	Induction Date
PD	Production Date
N	Production Load Norm
L	Direct Civilian Man-Hours Incurred
L$	Direct Labor Costs Incurred
M$	Direct Material Costs Incurred
PE	Production Expense Applied
GA	General and Administrative Expense Applied
NC	Total Cost NIF Work Completed
SC	Total Statistical Costs
TC	Grand Total NIF and Statistical Costs

APPENDIX B

Calculation of the Index of Standard Hours

The index of standard hours for a job is calculated as follows.

For any given job its induction and production date are used to establish the period of time during which that job is in the shop. The standard hours produced and the direct man-hours incurred in the production of those standard hours in the 953 shop (aircraft assembly) are found for the same period of time. The ratio of these standard hours to direct man-hours yield a measure of the efficiency of the 953 shop during the period of time in question. The direct man-hours incurred on the job are then weighted by this measure of efficiency to produce an index of the standard hours produced in the accomplishment of the job.

This approach is subject to several drawbacks as is the construction of any index. However, since the vast majority of the work in the 953 shop is on F4-B to N conversions, and since the major contribution of direct man-hours to any B to N job is from the 953 shop; this procedure seems defensible.

APPENDIX C

Conditional Estimation and Hypothesis Testing With Perfect Multicollinearity[1]

Suppose $y = z\delta + \varepsilon$ (A.1)

where $\underset{(nxK+3)}{z} = [X, P], X = [X_1, X_2], X_2 = [W, (1-W)]$

$\underset{(1xK+3)}{\delta^t} = [\beta_1^t, \beta_{21}, \beta_{22}, \gamma]$

Let $P = [W, (1-W)] \rho$

where $\rho = \begin{bmatrix} 1 \\ \rho_1 \end{bmatrix}$

Thus we have a problem of perfect multicollinearity since P is just a linear combination of W and (1-W). This is essentially the problem relating the dummy variables associated with WIPICS and the material price index.

If the problem is reformulated as Theil suggests then:

$$Y - P\gamma = X\beta + \varepsilon \quad \text{(A.2)}$$

and a conditional estimator may be derived as

---

[1]This Appendix follows the exposition of Theil [10, pp. 147-155].

$$\tilde{\beta} = (X^t X)^{-1} X^t (Y - P_Y) \tag{A.3}$$

$$= \hat{\beta} - (X^t X)^{-1} X^t P_Y$$

where $\hat{\beta}$ is the least square estimator which would be derived by merely dropping P from the regression equation. That is $\hat{\beta}$ is the estimator reported in (4.8).

But

$$P = X_2 \rho = X \begin{bmatrix} 0 \\ \rho \end{bmatrix} \tag{A.4}$$

therefore

$$\tilde{\beta} = \hat{\beta} - (X^t X)^{-1} X^t X \begin{bmatrix} 0 \\ \rho \end{bmatrix} \gamma$$

$$= \hat{\beta} - \begin{bmatrix} 0 \\ \rho\gamma \end{bmatrix} \tag{A.5}$$

$$= \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_{21} - \gamma \\ \hat{\beta}_{22} - \rho_1\gamma \end{bmatrix}$$

Also notice that

$$\text{var}(\tilde{\beta}) = \text{var}(\hat{\beta}) = \sigma^2 (X^t X)^{-1} \tag{A.6}$$

and it can be estimated by

$$\widehat{\text{var}\ (\tilde{\beta})} = S^2 (X^t X)^{-1} \tag{A.7}$$

Thus a conditional estimator for β, $\tilde{\beta}$, may be derived directly from (A.5) given values for $\hat{\beta}$, γ, and ρ_1. Also the estimated variance of the conditional estimator does not depend on the values of γ or ρ_1.

Finally to test the hypothesis $H_0 : \beta_{21} - \beta_{22} \leqq 0$ against the alternative hypothesis $H_a : \beta_{21} - \beta_{22} > 0$ we need only calculate:

$$t = \alpha^t \tilde{\beta} \Big/ \sqrt{S^2 \alpha^1 (X^t X)^{-1} \alpha} \tag{A.8}$$

where

$$\alpha^t = [\emptyset, 1, -1] \quad \text{and}$$

$$\alpha^t \tilde{\beta} = \hat{\beta}_{21} - \gamma - \hat{\beta}_{22} + \rho_1 \gamma$$

The denominator of (A.8) is just the estimated standard deviation of $\alpha^t \hat{\beta}$. This is the basis for the sensitivity analysis reported in Section 5.

THE EFFECTS OF RENEWAL PROCESSES UPON STOCHASTIC RELIABILITY MODELS

by

Capt Louis A. Dugas, Jr.
Capt David H. Hartmann
Lt Col Joseph E. Boyett, Jr.

Air Force Institute of Technology
Wright-Patterson AFB, OH 45433

Logistics policies typically assume that failure times are generated from a Poisson process. This assumption is usually caveated as being popular or a necessary simplifation for mathematical manipulation. In addition, major assemblies undergoing extensive repair or overhaul are themselves assumed to be brand new upon renewal. The influence of arbitrary maintenance policies upon the observed reliability of assemblies has not been thoroughly explored at present. This research uses a discrete event probabilistic simulation model to examine the impact of five realistic maintenance policies upon nine different assemblies. Although the study is exploratory and the conclusions are not final, indications are that the hazard function for an assembly does indeed pass through transient conditions when approaching a steady state. Maintenance policies established as a function of an assembly's hazard function must then change over time in order to move with the assembly's changing hazard function.

This paper is not included in the proceedings because of its length. It is available as AD #A030297 from the Defense Documentation Center, Cameron Station, Alexandria, VA 22314.

SOME GENERIC PROPERTIES OF A LOGIC MODEL FOR ANALYZING HARDWARE MAINTENANCE AND DESIGN CONCEPTS*

James T. Wong
William L. Andre

US Army Air Mobility Research
& Development Laboratory
Ames Research Center
Moffett Field, CA 94035

SUMMARY

A mathematical structure for diagnostic logic modeling was formulated, which allows the intrinsic properties of a complex Logic Model to be studied in an abstract setting. As a result, it was found that a loop-free Logic Model is a partially ordered set and that every permutation of the elements in the terminal set of a finite partially ordered set S partitions S into disjoint subsets. Based on these results, it was deduced that the minimum number of test points required for conclusive detection of malfunctioning components for a loop-free system is equal to the number of elements in the terminal set; this set constitutes the optimal choice for test points. Also, it was established, for each permutation of the elements in the terminal set, a relative failure probability measure. Based on this probability measure, an optimal diagnostic strategy was defined in accordance with Bellman's Principle of Optimality. Finally, for the purpose of illustration, some examples are given.

INTRODUCTION

Techniques used to analyze maintenance characteristics have been frequently subjective and difficult to determine at the conceptual and development stages of an equipment. This is largely due to the lack of analytical methods which allow the functional relationships between the hardware components or parts to be easily understood. Studies (1), (2), conducted by the US Army Air Mobility R&D Laboratory in reliability and maintainability technology, address this problem area.

The studies, entitled "Maintenance Logic Model Analysis Feasibility Investigation" and "LOGMOD Diagnostic Procedures for Mechanical and Electrical Equipment," address the feasibility of the computer generation

and manipulation of a Logic Model Algorithm for use as a design and maintenance tool. Included are diagnostic procedures for assessing troubleshooting and maintenance strategies associated with hardware design.

Automatic data processing techniques are used to generate an algorithm, which accepts engineering functional design data and organizes these data into a structure of least dependent to most dependent hardware elements. The structure, called the Logic Model (Diagram), is precisely a schematic representation of the intrinsic functional relationships between the elements of the equipment. The Logic Model was developed to provide improved troubleshooting analysis and evaluation of hardware design concepts for maintenance analysis purposes.

The Logic Model concept is considered an engineering innovation. Therefore, a well established mathematical basis for this concept should be formulated. This then allows the intrinsic properties of a complex Logic Model to be studied in an abstract setting. Some preliminary results in this regard have been published (3). The purpose here is to present a mathematical development of the generic concepts.

PARTIALLY ORDERED SETS

Well established mathematical concepts form the basis upon which the techniques discussed here are based.

Generally speaking, a partially ordered set P is a system consisting of a set S and a relation $\sim$ satisfying the following postulates:

1. $a \sim b$ and $b \sim a$ hold if and only if $a = b$, $a, b \in S$.
2. $a \sim b$ and $b \sim c$ imply $a \sim c$, $a, b, c \in S$.

If S contains only a finite number of elements, then P is called a finite partially ordered set. In this paper, we are concerned with finite partially ordered sets. For the purpose of illustration, the following are simple examples of partially ordered sets:

A. Let $S = \{1, 0, -4, 10\}$ and the relation $\sim$ be the usual "greater than or equal to" binary relation. The system is a finite partially ordered set, for obviously the set S together with the given relation satisfies the two postulates above.

B. Let S be the set of all real numbers in the unit interval with the relation as that of example A. Then the system is a partially ordered set, but not finite.

C. Let S be the set of all letters in the word "relation" and the relation is defined by the arrangement: relation, that is, $e \sim 1$ (read e *precedes* 1); this is a partially ordered set.

However, if we let S be the set of all lines on a plane and the relation be "parallel;" the system is *not* a partially ordered set even though it satisfies the transitive property 2 but not the asymmetric requirement 1.

BASIC ELEMENTS OF THE LOGIC MODEL

The following are the basic elements with which the logic model is constructed. For a more detailed discussion the reader is referred to the published reports (1) and (2).

Component - A physical component of an equipment.

Event - A measurable or observable quantity.

Functional entity - A component or an event.

Dependence - A functional relationship between two functional entities.

Input or independent event - An event for which there exists no functional entity on which the event is dependent.

Terminal event - An event upon which no functional entity depends.

Dependence chain - A collection of functional entities in which every two of its entites are dependent.

Loop - A closed dependence chain.

SOME INTRINSIC PROPERTIES

In this section we shall state and prove some fundamental characteristics of the Logic Model. The mathematical development of the model is an elementary exercise; for it merely translates the concepts into an algebraic structure. This mathematical structure forms a basis, from which some interesting results, that are useful for maintenance analysis, are deduced.

Lemma 1. A loop-free functional logic model is a partially ordered set.

Proof: Let S be the set of all functional entities of the logic model and the functional dependence, denoted by $<$, be the relation on S. Then, for $x_o, y_o \varepsilon S$, $x_o < y_o$ and $y_o < x_o$, the logic model input data requirements ensure that there exist $x_1, x_2, x_3, \ldots, x_m$ and $y_1, y_2, y_3, \ldots, y_n$ in S such that the following dependency chains can be realized:

$$x_o < x_1 < x_2 \ldots < x_m < y_o$$

and

$$y_o < y_1 < y_2 \ldots < y_n < x_o$$

So, concatenation of the two chains gives

$$x_o < x_1 < \ldots < x_m < y_o < y_1 < \ldots < y_n < x_o$$

Now if there exist two distinct elements in the set

$$A \equiv \{x_i, y_j \mid i = 0,1,2,\ldots,m;\ j = 0,1,2,\ldots,n\},$$

then the chain is closed; hence, it is a loop which contradicts the loop-free assumption. Therefore, all elements of A are identical. It remains to be shown the transitive property holds. Let x_o, y_o, and z_o be elements in S such that $x_o < y_o$ and $y_o < z_o$. Using the similar technique as before, one obtains $x_o < z_o$.

Lemma 2. For every finite partially ordered set, the terminal set is not empty.

Proof: Let S be a finite partially ordered set and T be the terminal set of S. To prove Lemma 2, assume the contrary. Let $a_1 \varepsilon S$. Then $a_1 \notin T$ implies there exists an element $a_2 \varepsilon S$ such that $a_2 < a_1$. Now $a_2 \notin T$ implies there is $a_3 \varepsilon S$ such that $a_3 < a_2$. Thus, continuing the process inductively, one obtains a dependency chain, $a_k < a_{k-1} < \ldots < a_2 < a_1$, where k is a positive integer and is less than or equal to the number of elements in S. But $a_k \neq a_i$, for all $i = 1,2,\ldots,k-1$, since S is a partially ordered set. Therefore, a_k is a terminal point, by definition. Hence, the conclusion follows.

Lemma 3. Each permutation of the elements in the terminal set T of a finite partially ordered set S partitions S into disjoint subsets whose union is precisely S.

Proof: By Lemma 2, T contains k elements, for some positive integer k. Now, let the ordered k-tuples $(a_1, a_2, a_3, \ldots, a_k)$ be a permutation of the elements of T. Also, we define

$$H_{a_o} \equiv \Phi ,$$

and

$$H_{a_i} \equiv \left\{ x\varepsilon S - \bigcup_{\ell=0}^{i-1} H_{a_\ell} \middle| a_i < x \right\},\ i = 1,2,\ldots,k.$$

Note that H_{a_i} are subsets of S, and, if $k = 1$, the Lemma 3 conclusion is obvious. On the other hand, let $m,n\varepsilon\{1,2,3,\ldots,k\}$ such that $m \neq n$ and

$$H_{a_m} \cap H_{a_n} \neq \Phi.$$

So, let $y\varepsilon H_{a_m} \cap H_{a_n}$. Then $y\varepsilon H_{a_m}$ and $y\varepsilon H_{a_n}$. Without loss of generality, assume $m < n$. This implies that there exists an integer $j\varepsilon\{1,2,\ldots,k-1\} \ni n = m + j$. By definition,

$$H_{a_n} = \left\{ x\varepsilon S - \bigcup_{i=0}^{n-1} H_{a_i} \middle| a_n < x \right\}.$$

It follows that $y\varepsilon H_{a_n}$ implies

$$y \notin \bigcup_{i=0}^{n-1} H_{a_i} = \bigcup_{i=0}^{m+j-1} H_{a_i} \supset H_{a_m}$$

Therefore, $y \notin H_{a_m}$, which contradicts the assumption that H_{a_n} and H_{a_m} have nonempty intersection. Hence, these subsets are all disjoint. It remains to be shown that their union is equal to S. Suppose $x\varepsilon S$. Then either $x\varepsilon T$ or $x\varepsilon S-T$. If $x\varepsilon T$, then

$$x\varepsilon \bigcup_{i=1}^{k} H_{a_i}.$$

If $x\varepsilon S-T$, then $x\notin T$. $x\notin T$ implies there exists a dependency chain which can be described as

$$x < x_1 < x_2 \ldots < x_{q-1} < a_q,$$

for some integer $q\varepsilon\{1,2,\ldots,k\}$. Therefore, $x\varepsilon H_{a_q}$. Also, since the union of subsets is again a subset, we have

$$S = \bigcup_{i=1}^{k} H_{a_i} ,$$

which completes the proof of Lemma 3.

Having these results at our disposal, it is a simple exercise to deduce the following consequence pertinent to diagnostics as applies to analyzing hardware maintenance and design concepts.

Theorem 1. The minimum number of test points required for conclusive detection of malfunctioning components for a loop-free functional system is equal to the number of elements in the terminal set. This set constitutes the optimal choice for test points.

Proof: Lemma 3 implies that any malfunctioning entity in the set S is detectable by testing the terminal points. Also, malfunctioning of any terminal point is not detectable by testing the other terminal points or events.

Before we continue, let us examine the content of Theorem 1. The apparent drawback of Theorem 1 is its limitation to systems not containing a functional dependency loop. For system containing loop, the conclusion of the theorem is, in general, false unless the loop can be degenerated into a functional entity which is not a terminal point. For systems containing a functional loop whose degeneration gives rise to a terminal point, a remedy for this problem is to add an additional test point on the loop; for then the theorem is applicable to the modified system.

Although Theorem 1 provides some quantitative information for conclusive detection of system malfunctioning, no information can be deduced from it concerning some "optimal" testing sequence on the terminal events. In what follows, an attempt is made to analyze a

test procedure for the case in which the terminal set contains two or more elements. In particular, we propose to determine a diagnostic testing sequence (or strategy) for conclusive detection of system malfunctioning such that the expected expenditure of resources is minimum.

For the analysis of this problem, we observe that each functional entity of a logic model relating to a hardware component or measurable item is normally considered as a potential source of failure, and therefore a measure of failure is required to determine the relative failure probability associated with the affected functional entities in the model. With this in mind, we state and prove the following statement:

Theorem 2. Each permutation of the elements in the terminal set induces a relative failure probability on the corresponding partition of a finite partially ordered set.

Proof: As before, let T be the terminal set containing k elements, and $(a_1, a_2, a_3, \ldots, a_k)$ be a permutation of the elements in T. Also, let

$$H_A \equiv \{H_{a_i}\},$$

where $i = 1,2,3,\ldots,k$, and H_{a_i} defined as before. Now, for each $x \in H_{a_i}$, let p_x be the failure probability of the functional entity x. Then the relative failure probability P_{a_i} associated with the set H_{a_i} is defined to be

$$P_{a_i} \equiv \sum_{x \in H_{a_i}} p_x \Big/ \sum_{y \in S} p_y .$$

Now, we define a function μ_A on H_A by the formula

$$\mu_A(H_{a_i}) = P_{a_i}$$

In view of Lemma 3, we have

$$\mu_A(H_A) = 1 ,$$

and obviously,

$$0 \leq \mu_A(E) \leq 1 ,$$

for any subset E of the power set of H_A. Hence, the assertion holds, since the function μ_A so constructed is a relative failure probability defined on the sample space.

The failure probability μ_A constructed in the proof of Theorem 2 depends on the permutation A, and $\mu_A(H_{a_i})$ is the relative probability of observing a failure belonging to H_{a_i} by testing the test point a_i. Therefore, the failure probability is a function of the testing sequence and the component failure probabilities.

Now we are in a position to define an optimal strategy for conclusive detection of system malfunctioning. To this end, we cite Bellman's Principle of Optimality from literature (4) as follows:

> An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.

Also, we define: An admissible strategy A is that strategy such that the expected expenditure (cost or time) incurred due to applying A to a given functional system in minimum. Symbolically,

$$\min_{A \in \Lambda} \sum_{i=1}^{k} U_{a_i} P_{a_i}$$

where $A = (a_1, a_2, a_3, \ldots, a_k)$

a_i = terminal point

Λ = the set of all possible strategies

U_{a_i} = the expenditure associated with the test point a_i

P_{a_i} = the relative failure probability of all the dependency chains terminated at a_i

An optimal strategy is an admissible strategy which obeys the Principle of Optimality.

EXAMPLE MODELS

A pictorial representation of a logic model can be depicted by using arrows and points. An arrow is used to indicate functional or physical dependency between two functional entities, and a point indicates a particular entity. A working logic model can be complex containing several hundred functional entities and dependencies, however, only a simple model is needed for illustration of the concepts involved.

Figure 1 shows an example of an intermediate loop in a model which does not include any terminal points. This is a three point loop. In general, any loop can involve two or more points. In Figure 2, a terminal loop is illustrated and involves four points. If the loop were to degenerate to a point the point would be a terminal event for the model. If terminal loops exist in a design it will often mean that troubleshooting and diagnosis of the hardware when malfunctioning can be challenging if not impossible without replacing the entire piece of equipment.

A third example is illustrated by Figure 3 to illustrate a sample calculation for a loop-free system. The numerical values at each point are used to represent a measure of failure probability for the functional entity represented by the point. The three terminal points

which exist are labeled simply as 1, 2, and 3, respectively. For this case then there are a total of 3! strategies possible for executing a test procedure to detect a malfunctional entity. These are:

STRATEGY	TERMINAL POINT TEST SEQUENCE
A	$2 \rightarrow 1 \rightarrow 3$
B	$1 \rightarrow 2 \rightarrow 3$
C	$3 \rightarrow 1 \rightarrow 2$
D	$3 \rightarrow 2 \rightarrow 1$
E	$2 \rightarrow 3 \rightarrow 1$
F	$1 \rightarrow 3 \rightarrow 2$

For strategy A the partitioned sets of elements which exist are indicated by the dashed lines, Figure 4. This partitioning would change for each strategy and therefore the corresponding probability for detection of a malfunction at each test point will change with each strategy.

The probabilities of detection and expected expenditures for each strategy (A through F) are presented in Table 1 and Table 2. The probability of detection at each test point is listed in the strategy column and the entries within the table show the expected expenditure in terms of cost and time, respectively, in the two tables. The numerical values associated with expenditures at each test point are given at the bottom of each table. Note that in Table 1 there are three strategies which yield a minimum value of 1.08. These are considered admissible strategies as previously discussed. However, only strategy F is considered optimal. Therefore, on a cost basis, this is the best choice for

testing. However, on a time basis, strategy A is optimal as shown in Table 2. This is an interesting situation. A time-cost tradeoff must be made if a single best strategy is to be chosen.

Figure 5 plots the values of expected time versus expected cost for each strategy and indicates the overall preferred strategy to be B.

This type of analysis is normally performed by a computer for models of any significant size where the number of strategies grows factorially with the number of test points.

CONCLUSIONS

The logic modeling concept is considered to be an engineering innovation. In this paper, the mathematical structure for a class of logic models was established, and some of its generic properties, useful for maintenance analysis, were deduced. Specifically, we make the following concluding remarks:

1. A loop-free functional logic model is a partially ordered set.

2. For every finite partially ordered set, the terminal set is not empty.

3. Each permutation of the elements in the terminal set T of a finite partially ordered set S partitions S into disjoint subsets whose union is precisely S.

4. The minimum number of test points required for conclusive detection of malfunctioning components of a loop-free functional system is equal to the number of elements in the terminal set. This set constitutes the optimal choice for test points.

5. The conclusion of statement 4 is valid for systems containing loops only if the loop can be degenerated into a functional entity which

is not a terminal point. If the degenerated entity is a terminal point, then statement 4 is applicable if the system is modified to allow an additional test point in the loop.

6. Each permutation of the elements in the terminal set induces a relative failure probability for the corresponding partition of S.

7. Based on statement 6, an admissible diagnostic strategy A is defined to be a strategy such that the expected expenditure (cost and/or time) incurred due to applying A to a given functional system is a minimum.

8. An optimal strategy is an admissible strategy which obeys the Principle of Optimality.

9. An optimal strategy for a given system need not be unique.

REFERENCES

1. DePaul, Ralph A., Jr., Maintenance Logic Model Analysis Feasibility Investigation, Advanced Systems Research Office Report, U.S. Army Air Mobility R&D Laboratory, Moffett Field, California, Nov. 1974.
2. DePaul, Ralph A., Jr., and Dingle, Larry A., LOGMOD Diagnostic Procedures for Mechanical and Electrical Equipment, Advanced Systems Research Office Report, U.S. Army Air Mobility R&D Laboratory, Moffett Field, California, June 1975.
3. Andre, William L., and Wong, James T., Mathematical Structure and Optimal Diagnostic Technique for Maintenance Analysis, Presented at and published in the *Proceedings of the 36th Military Operations Research Society*, Dec. 1975.
4. Bellman, R. E., and Dreyfus, S. E., Applied Dynamic Programming, Princeton University Press, 1962.

TABLE 1.- EXAMPLE OF COST COMPUTATION

Strategy	Expected cost 1st T.P.	Exp. cost 1st and 2nd T.P.	Exp. cost 1st and 2nd and 3rd T.P.
A 2 → 1 → 3 .36 .32 .32	.72	1.04	1.36
B 1 → 2 → 3 .40 .28 .32	.40	.96	1.28
C 3 → 1 → 2 .84 .08 .08	.84	.92	1.08
D 3 → 2 → 1 .84 .08 .08	.84	1.00	1.08
E 2 → 3 → 1 .36 .56 .08	.72	1.28	1.36
F 1 → 3 → 2 .40 .52 .08	.40	.92	1.08

For: $C_1 = 1$, $C_2 = 2$, $C_3 = 1$

TABLE 2.- EXAMPLE OF TIME COMPUTATION

Strategy	Expected time 1st T.P.	Exp. time 1st and 2nd T.P.	Exp. time 1st and 2nd and 3rd T.P.
A 2 → 1 → 3 .36 .32 .32	.36	1.00	1.96
B 1 → 2 → 3 .40 .28 .32	.80	1.08	2.04
C 3 → 1 → 2 .84 .08 .08	2.52	2.68	2.76
D 3 → 2 → 1 .84 .08 .08	2.52	2.60	2.76
E 2 → 3 → 1 .36 .56 .08	.36	2.04	2.20
F 1 → 3 → 2 .40 .52 .08	.80	2.36	2.44

For: $T_1 = 2$, $T_2 = 1$, $T_3 = 3$

FIGURE CAPTIONS

Figure 1.- Model with intermediate loop.

Figure 2.- Model with terminal loop.

Figure 3.- Model for sample calculation.

Figure 4.- Model for sample calculation (strategy A).

Figure 5.- Combined cost/time plot.

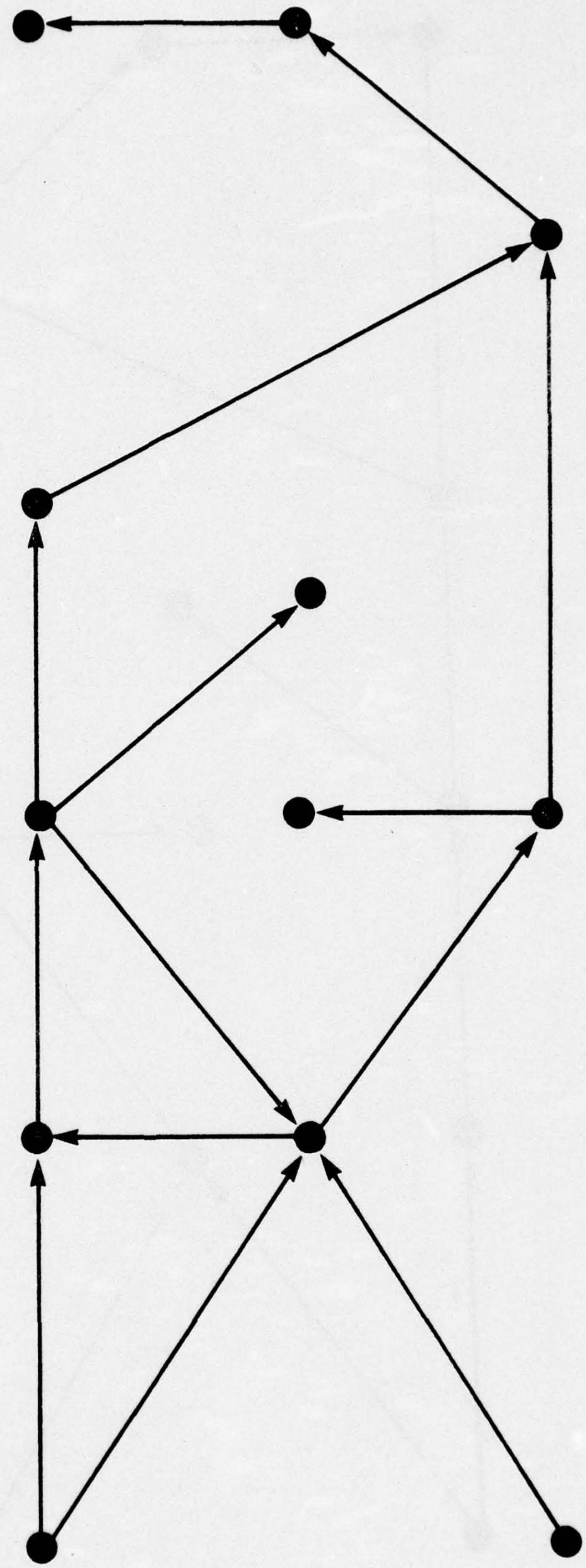


Fig. 1

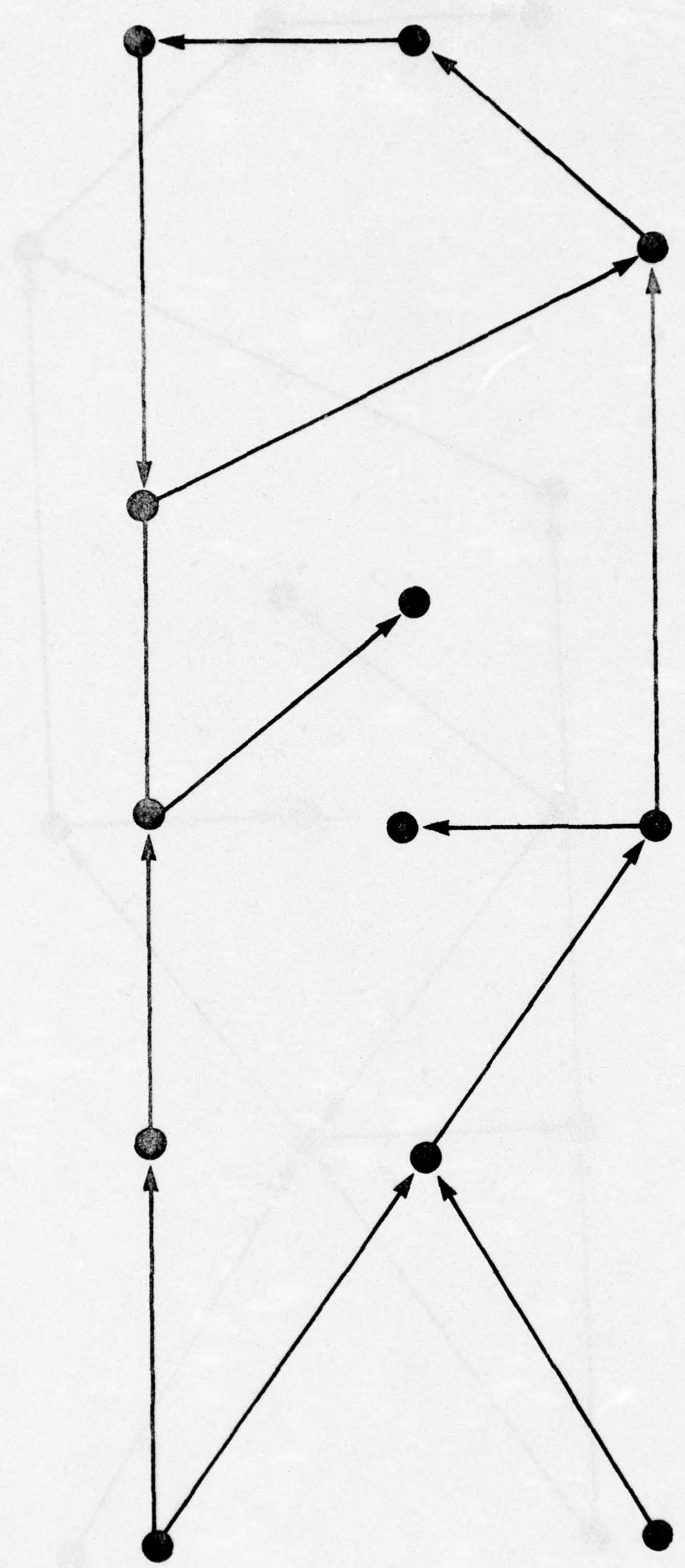


Fig. 2

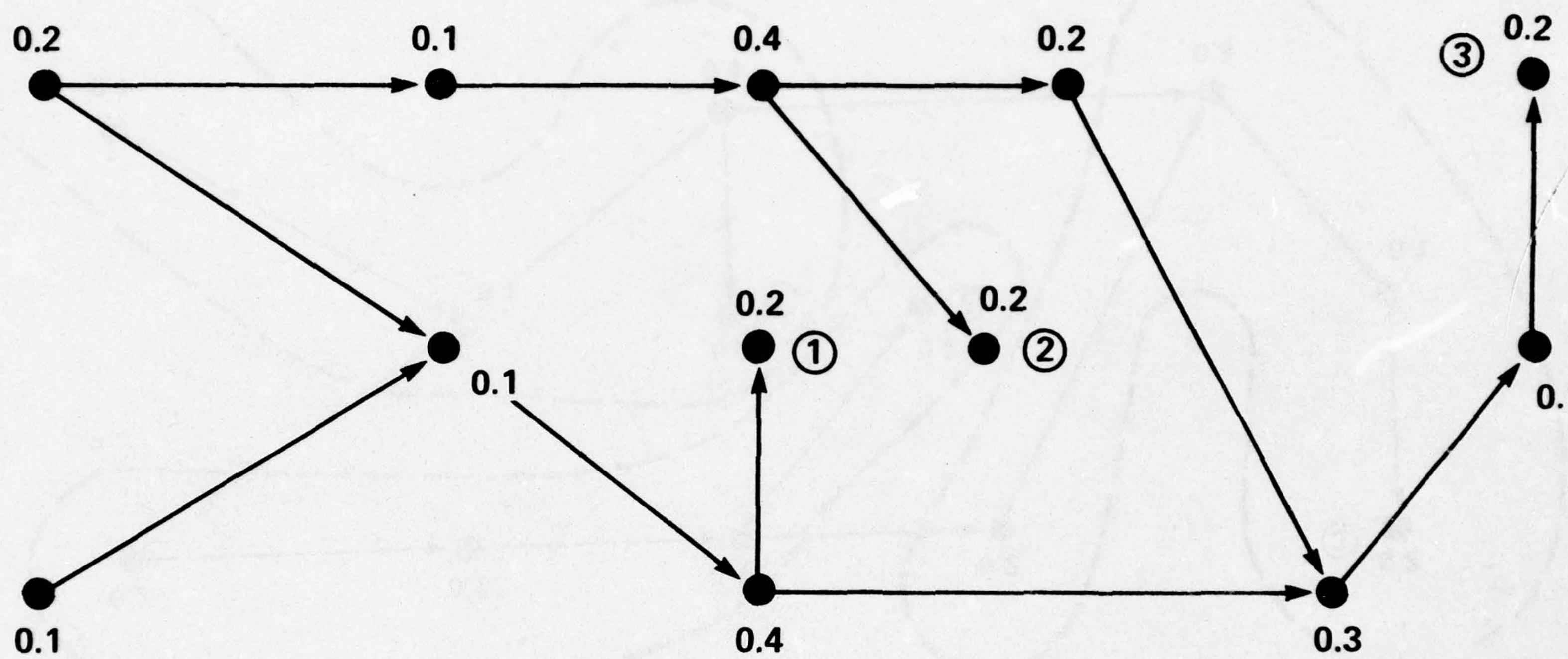
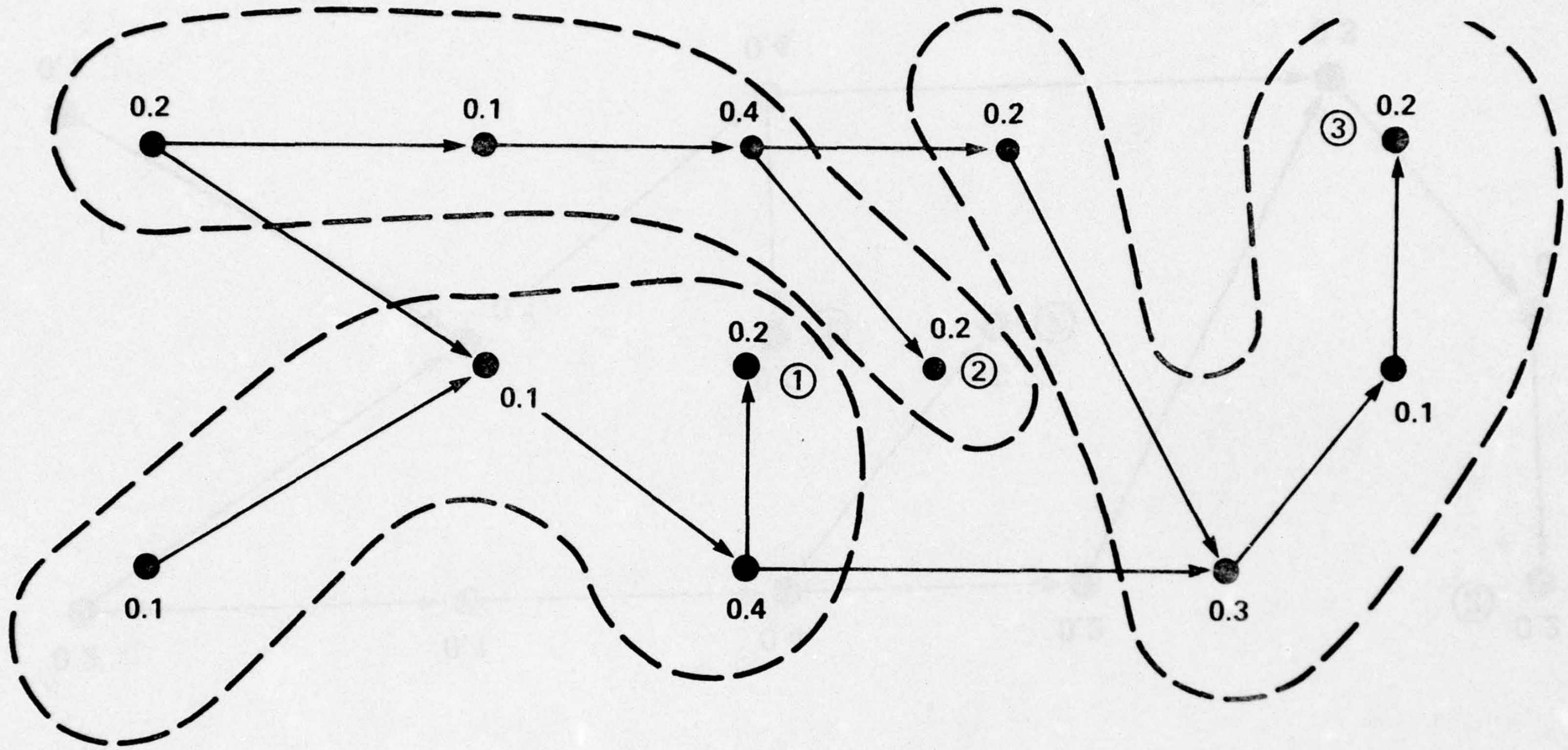


Fig. 3



STRATEGY A: ② → ① → ③

Fig. 4

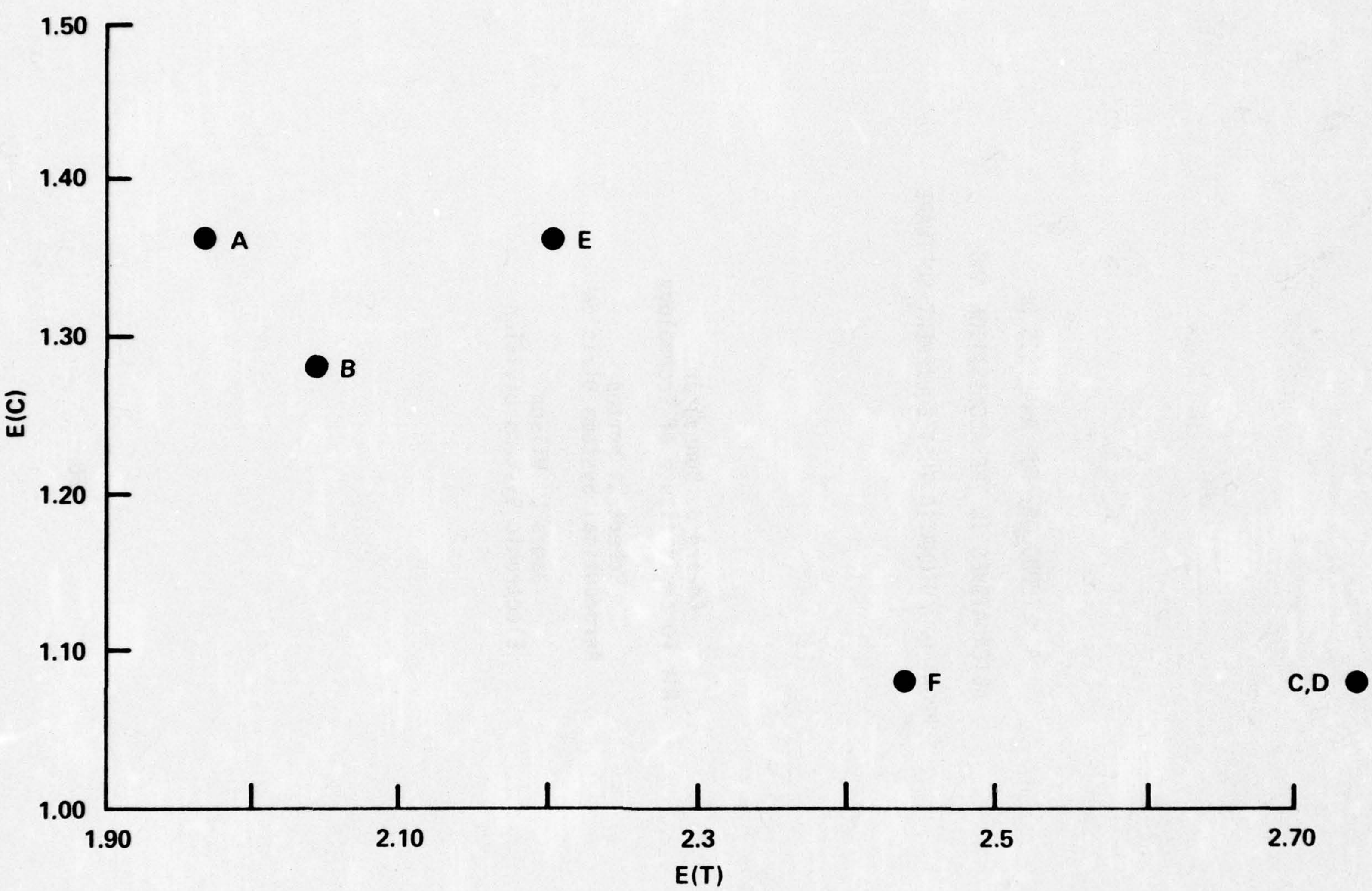


Fig. 5

A METHODOLOGY FOR ANALYSIS OF

ALTERNATIVES IN THE ACQUISITION AND

SUPPORT OF AUTOMATIC TEST EQUIPMENT SOFTWARE

Edward J. Dunne, Jr.
Air Force Institute of Technology

Robert W. Morton
Aeronautical Systems Division

James L. Wilson
Electronic Systems Division

INTRODUCTION

The focus of this paper is the support of the software element of automatic test equipment systems, or ATE systems. We define support to be the change and correction of discrepancies which is a constant requirement for ATE test software. Test software is computer programs in the ATE hardware which cause it to measure critical parameters of the item being tested.

Over the last few years, an increasing number of Air Force organizations have taken positive steps to improve the acquisition and support of the software used with ATE. Conferences have been held. Studies have been conducted. New organizations have been formed, and regulations have been published. Without question, important progress has been made in solving some problems and in bringing others into better focus. But many of the resulting recommendations and actions have dealt in generalities or absolutes. The recommendations of major studies in this area tend to imply that specific courses of action are applicable in all situations. The usual recommendations are to buy more documentation, more training, standardize the programming language, and expand the organic support capability.

The genesis of this research was our belief that "acquisition managers" have had little visibility into the long term impacts of their decisions in the ATE software area. They are driven very strongly by near term costs because of the way the Department of Defense acquisition "system" works and because there was no data available on long term effects. This research doesn't alleviate the first constraint, but does provide a tool which will help in the second area.

THE APPROACH

Our premises are (1) that ATE software support (the cost and responsiveness characteristics) is influenced by factors subject to decisions made during the acquisition phase (the ATE software acquisition management interfaces are shown in Figure 1), and (2) that the support of ATE software

(the activities of making necessary corrections/modifications to it) can be modeled as a set of interrelated processes with definable attributes, e.g. timeliness, cost. With these premises, the general approach was first to hypothesize influencing factors which were investigated, analyzed, and tested. Then, the Air Force support process was modeled to include these influencing factors which come from early acquisition decisions.

There are some aspects of the problem which seem obvious. Clearly, the decision to provide organic, contractual, or mixed organic and contract support of the software is crucial. Most people also recognize that the decisions on test program documentation will affect the long term support capability. The problems which need resolution are those related to the questions of how much, and under what circumstances. In the case of ATE software support, the questions do not have unequivocal or simple answers. The interacting policies, decisions, and processes related to ATE software maintenance form a complex support system. It is this complexity which is the heart of the problem.

System modeling is one approach to dealing with complexity. Where narrative and analytical models are so complicated that solution of the model rivals the difficulty of less formal approaches, simulation techniques offer an alternative. This research has been directed toward the development of a general simulation model which allows managers associated with ATE software acquisition and support to investigate the probable outcomes of adopting one of several alternative courses of action.

INFLUENCING FACTORS DERIVED FROM ACQUISITION DECISIONS

There were three factors which derive from acquisition phase decisions which we hypothesized would affect ATE software support. A major portion of the research effort was devoted to deriving a relationship which would describe the effects which programming language level (L), the amount of documentation (D), and the amount of training programmers received (T) have on the time required for programmers to complete software support tasks. It was necessary to determine the form of the effect and to implement a mathematical description of it in the simulation model.

We hypothesized three levels for each of these influencing factors, as follows.

Language Levels

H = Standard Problem-Oriented Language

M = Nonstandard Problem-Oriented Language

L = Tester=Oriented Intermediate Language

Documentation Levels

H = Complete, for example Test Requirements Document

M = Intermediate, listings and flow charts

L = Minimum, only listings

Training Levels

H = four to six weeks training

M = two to three weeks training

L = one week or less training

Data was collected from Air Force programmers with ATE software support experience on how these different factor levels affected their "time to accomplish a service action on a typical job." Using this data, adjustments were derived using regression analysis for different conditions--a situation with a specific level of each factor. This LDT adjustment was incorporated into the support process model.

THE MODEL

This model was developed based upon extensive observation, interviews, and data gathering within the Air Force ATE software support systems at Ogden, Warner Robins, and Sacramento Air Logistic Centers. The model may be viewed as a network of events through which transactions (discrepancy reports, work orders, contract calls, corrected tapes, etc.) flow. The model treats each type of transaction as one form of the service demand.

Figure 2 shows the very basic schematic form of the model. Service demands enter the model on a random basis as determined by the Demand Creation and Classification Segment. After a routine (none-ECP) demand is created, it passes through the Initial Processing and Support Source Selection segment. Demands identified as ECP-related can be moved directly to contractor support. The support source selected for routine demands is based on several alternative support policies which may be specified by the model user. From this point on, the processes which any demand encounters are quite different depending upon the organic/contract decision. When contract support is selected, costs and time delays are incurred for preparation of a statement of work, funding approvals, procurement processing, contractor services, and technical order printing and distribution. If a demand is organically serviced, delays and costs are incurred for preparation of a work order, computer support, programmer efforts, technical order editing, negative preparation, printing, and distribution. Within the organic segment, there are a maximum of twenty programmers. Each programmer is modeled separately. Within this segment, task size is determined as an observation of random variable and the language, documentation, and training effects are included. A time dependent on-the-job training effect also can be used.

The model operates on a time base where the smallest increment is one hour, representing one working hour of an eight hour day, five day week. The user specifies the number of years the simulation runs and the starting points of the random number generators.

A large amount of data can be obtained from a model run. Data is provided in four classes: (1) model parameters used in the simulation run, (2) quarterly data, (3) annual statistics, and (4) end of simulation statistics. To illustrate annual statistics, the average number of demands in the system, the cost of organic programmer services, the cost of organic idle time, among others, are provided. At the end of the simulation, all cost and time data are summarized.

Viewed from this perspective, the operation of the model is extremely straightforward. It is within the modules and segments that the essential characteristics of the simulation are contained. A complete description of

this model and its computerized implementation may be found in a recent AFIT document which is Reference 1.

SO WHAT

The result of this research is a simulation model of the Air Force ATE software support process which incorporates the influences of factors whose "levels" are determined primarily in the presupport acquisition phase. This capability will permit acquisition managers to weigh the effects of the alternatives to the decisions they must make concerning the support system. Additionally, since the Air Force ATE software support process is modeled, the effects of changes within that process can be evaluated.

The limitations associated with this capability must also be understood. The Air Force ATE software support system was modelled, not any support system. The data used to specify the parameters and the random variables in the system model were thoroughly researched but were incomplete in some instances and of poor fidelity in some other cases. The sources of the data and specific cautions are included in Reference 1. Model validation has been limited to a "face validity" effort reviewing the functional form and sample outputs with experienced personnel within the Air Force ATE software support system.

We think this specific methodology might be used in the following applications related to initial acquisition.

(1) Investigating the desirability of developing a standard ATE programming language
(2) Selecting a language level for a new test system development
(3) Selecting a level of ATE software documentation for a particular program
(4) Selecting a training level for organic ATE programmers
(5) Selecting an ATE software support concept

Even after the initial acquisition decisions have been made, there are times when this methodology can be of assistance to Air Force managers.

(1) Investigating budgeting requirements for organic software support
(2) Manpower planning for organic ATE software support

(3) Investigating mixes of organic and contract support when manpower resources are limited.

From a broader perspective, we feel that the type of research reported here may have much expanded application than simply the Air Force ATE software support system. Modelling efforts aimed at developing understanding of the repair and support processes, especially the interactions between presupport decisions and the resulting processes, can help provide acquisition managers insight into probable future implications of decisions among alternatives being weighed today. These are the kinds of things which must be done if we, the Department of Defense, are ever going to do more than pay lip service to "life cycle costs."

EXAMPLE APPLICATIONS

Analysis of an Acquisition Problem

To demonstrate the application of the methodology to consideration of alternatives which might be of interest during the acquisition of ATE software, a sample problem was constructed and analyzed. The purpose of the problem was to demonstrate how the model could be used to provide information bearing on a decision to develop an organic capability and in selecting a documentation level. In the problem the following assumptions were made:

(1) The test programs are used with a family of intermediate level ATE used to test avionics Line Replaceable Units (LRU's).

(2) There are 350 test programs to be supported.

(3) The programming language is a problem-oriented language, but peculiar to this family of ATE.

(4) Funds are available for only a week, introductory training course.

The problem which is posed is two-fold. First, should organic or contractor support be adopted? And second, if organic support is reccommended, what level of documentation should be acquired? The contractor has provided the following information:

- --Delivery of the required compiler and the associated compiler documentation will cost $440,000.
- --Delivery of program listings alone will cost $500 per program.
- --Delivery of program listings and flow diagrams will cost $2,000 per program.
- --Delivery of Test Requirements Documents (TRD's) and Program listings will cost $10,000 per program.

To investigate this problem, a four-step procedure was used. First, the model parameters were set to establish the desired characteristics. Second, the model was run for four different situations--once for total contractor support, one time each for organic support of routine demands with low, medium, and high levels of documentation. Next, the model results were analyzed. As a final step, the support impacts were compared with the initial acquisition costs. As a note, the simulation was replicated three times for each of the four situations. This was done to determine the effect which the random processes in the model had on the data.

Table 1 shows the results in terms of the total support costs for eight years of operation for each policy. Statistical hypothesis tests allowed rejection of the null hypotheses that these costs were identical for any pair.

Adding the original documentation costs plus the compiler costs, the total costs are shown in Table 2. This calculation of total acquisition and support cost has been computed without considering the time value of money concept. Discounting was not used to keep the example simple. Wtih these calculations, the policies can now be ranked: Policy IV > Policy III > Policy I > Policy II.

Other information can be obtained from the simulation which should be considered along with the cost data. These data are those related to the responsiveness of the support. One potentially key parameter is the average time taken to satisfy a routine demand. This value is identified in Table 3 for each policy.

If a graph such as Figure 3 is used, one can visualize both cost and responsiveness aspects of each of the four policies. When examined only in terms of total costs, there is not a large difference in Policies I and II. Likewise, Policies II and III would lead to about the same average time to complete a routine demand. But, when cost and responsiveness are viewed together, Policy II seems to be the one which would be preferred. This, of course, is based on the assumption that the decision-maker places some positive value on a decrease in average service time per demand.

Analysis of Support Manpower Requirements

The second example was chosen to demonstrate one way in which the ATE Software Support Simulation Model can be used to investigate organic manpower assignment trade-offs. In this example all of the model parameters will be held constant except the number of programmers. By using the model, alternative manpower policies will be examined over a two-year period.

The first step taken in examining this problem was to run the simulation for two years with the number of programmers held constant at 20. This was done to obtain the quarterly printout of the average number busy in each quarter. Using this information, several alternative policies were hypothesized and examined.

A total of ten combinations of manpower assignments were simulated. Selected outputs of the simulation are shown in Table 4. The results here are for a single experiment for each policy. Were this an actual problem, some replication for each policy would be advised. From the data in Table 4, a graph was constructed similar to the one used with the first example. Here, average programmer costs per demand were used as a measure of the monetary impact of a policy. Figure 4 shows the ten combinations on a single graph. No attempt will be made to identify a best policy.

It can be seen that significant trade-offs do exist. Lower average service times can be obtained for the price of an increase in total manpower commitment and average cost per demand. In general terms, one can choose any one of several alternatives depending on the objectives which

one desired to achieve or depending upon existing constraints. Without the use of some tool, it is not clear how such trade-offs could otherwise be considered.

The two examples presented provide a means of investigating two hypothetical ATE software acquisition and support problems. The results of the simulations provide the reader a means to judge whether the model does provide insights which were not otherwise obvious.

REFERENCES

1. Wilson, James L. and Robert W. Morton, A Methodology for Analysis of Alternatives in the Acquisition and Support of Automatic Test Equipment Software, Air Force Institute of Technology Masters Theses GSM 74D-8, December, 1974.

2. User's Guide, ATE Software Support Simulation Model, Air Force Institute of Technology, Department of Systems Management, December 1974.

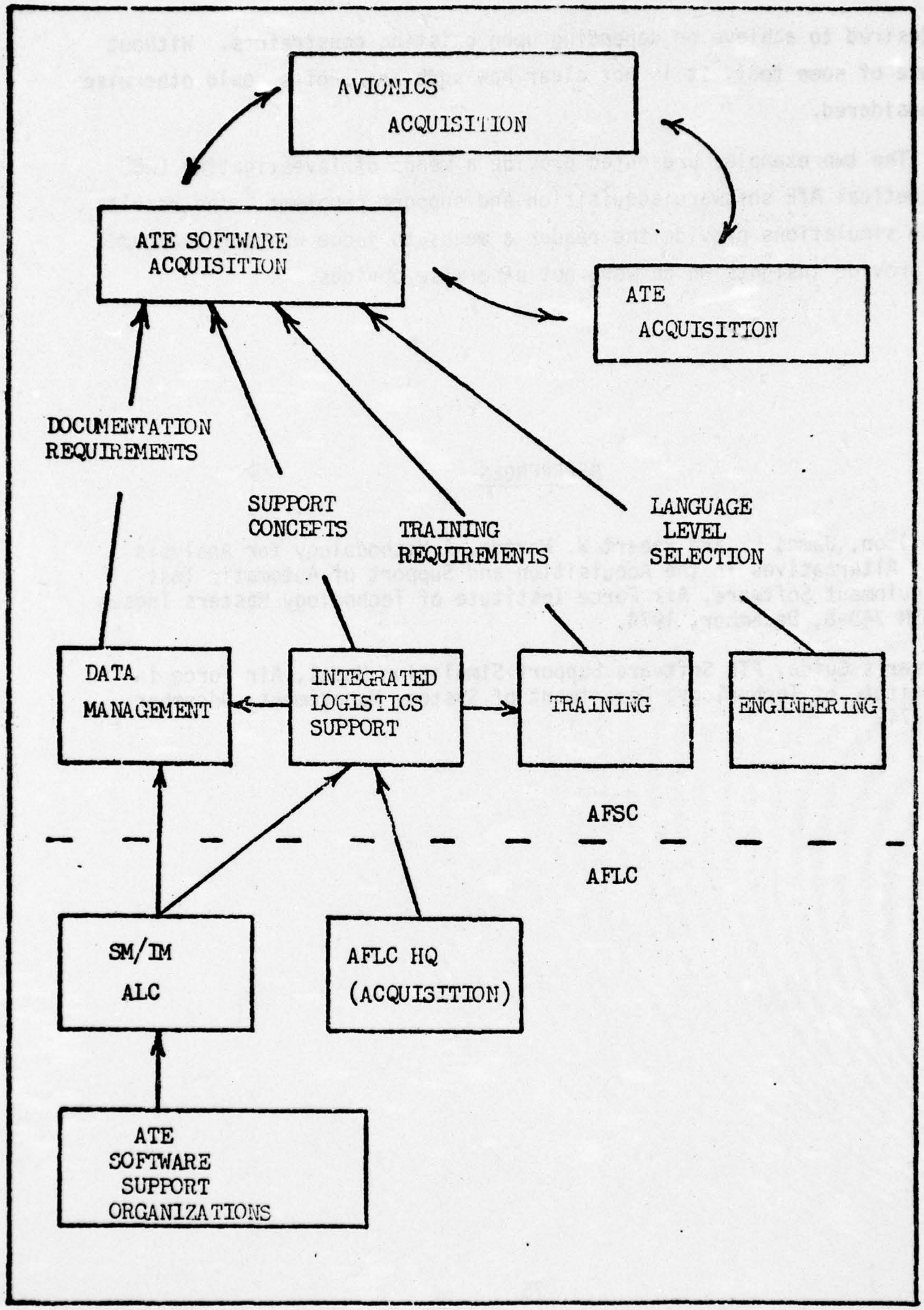


FIGURE 1

ATE Software Acquisition Management & Interfaces

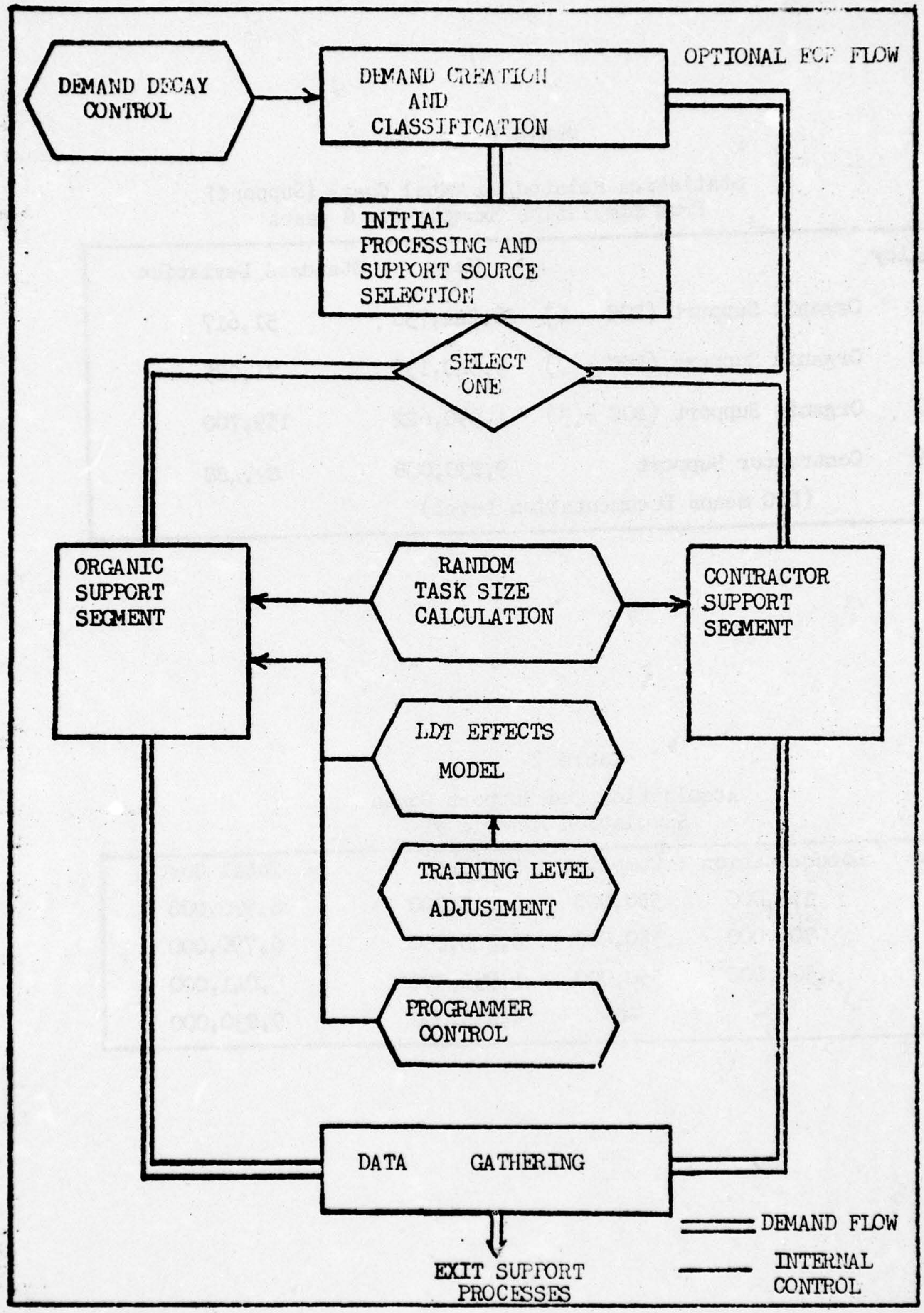


FIGURE 2

Simulation Model Form

Table 1

Statistics Related to Total Costs (Support)
From Simulation Example #1, 8 years

Policy		Mean	Standard Deviation
I	Organic Support (DOC = 1)	$6,244,736	51,617
II	Organic Support (DOC = 2)	5,540,136	71,085
III	Organic Support (DOC = 3)	4,990,622	159,700
IV	Contractor Support	9,830,028	87,488

(DOC means Documentation Level)

Table 2

Acquisition and Support Costs
Simulation Example #1

Policy:	Documentation +	Compiler +	Support =	Total Cost
I	175,000	550,000	6,245,000	6,970,000
II	700,000	550,000	5,540,000	6,790,000
III	3,500,000	550,000	4,991,000	9,041,000
IV	-0-	-0-	9,830,000	9,830,000

Table 3

Responsiveness Data Simulation Example #1

Policy	Average Time Per Demand (Routine only)
I	858 hours or 5.0 months
II	627 hours or 3.6 months
III	559 hours or 3.2 months
IV	789 hours or 4.5 months

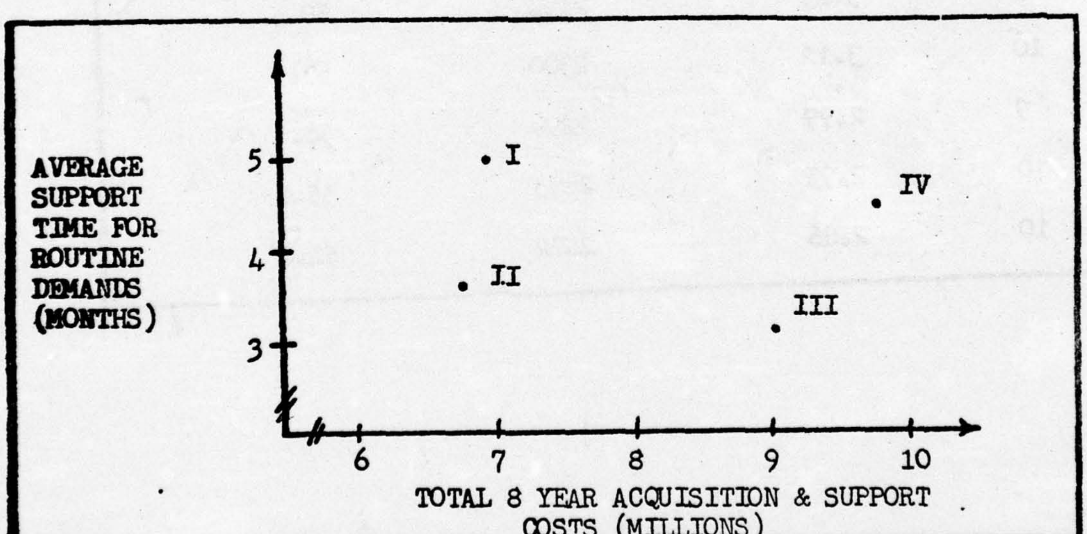


FIGURE 3

Policy Comparison Example #1

Table 4

Data for Simulation Example #2

Number of Programmers Yr 1	Yr 2	Avg. Time Routine Demands (Mos)	Avg. Cost per Demand	Avg. Number in System
12	9	4.64	2162	78.0
13	8	4.19	2167	72.0
13	9	4.06	2172	70.0
13	10	3.85	2202	67.0
14	8	3.42	2168	61.5
14	9	3.25	2196	59.5
14	10	3.12	2300	57.5
15	9	2.99	2299	56.0
15	10	2.95	2395	55.0
16	10	2.86	2472	54.0

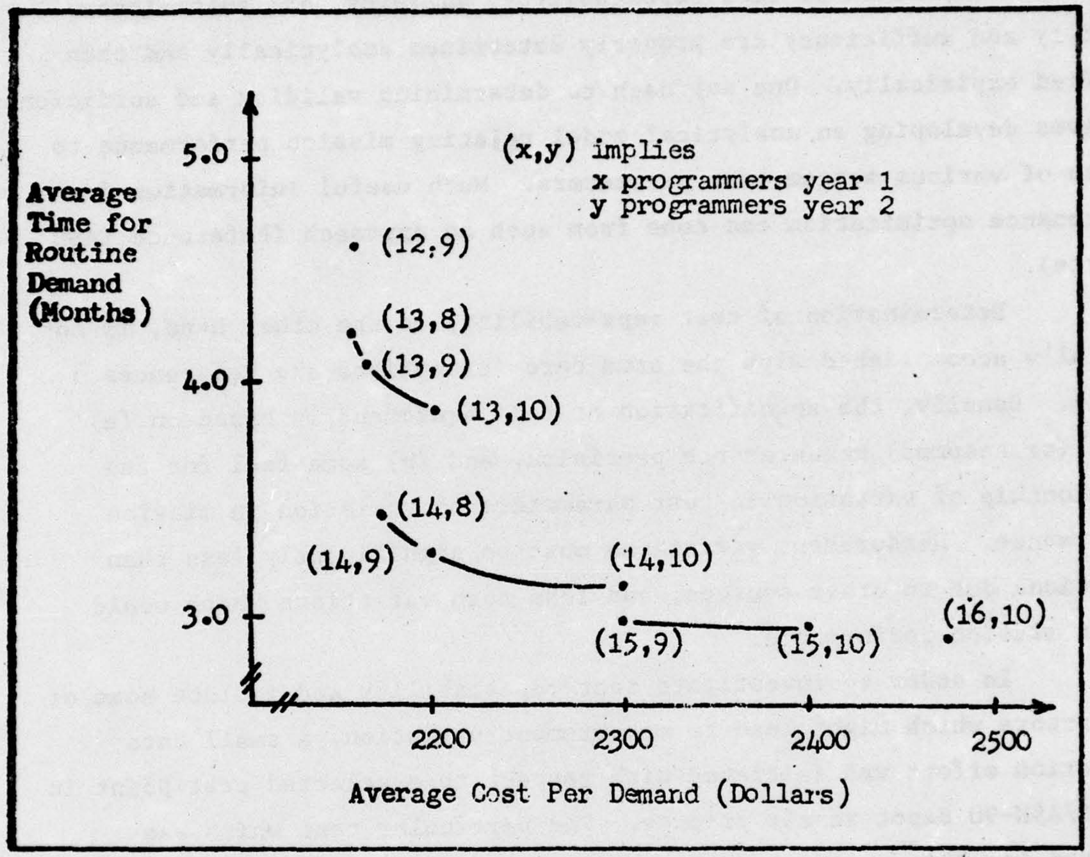


FIGURE 4

Policy Comparison Example #2

ASSESSING TEST REPEATABILITY

Girard Levy and Horace Ray

Battelle Columbus Laboratories, Columbus, Ohio

One of the general problems in optimizing maintenance decisions is the determination of test repeatability, validity, and sufficiency. Validity and sufficiency are properly determined analytically and then verified empirically. One approach to determining validity and sufficiency involves developing an analytical model relating mission performance to values of various system test parameters. Much useful information for maintenance optimization can come from such an approach (Reference TASC reports).

Determination of test repeatability, on the other hand, is not typically accomplished with the same care (exceptions are References 1 and 2). Usually, the specification of test equipment is based on (a) known (or assumed) measurements precision, and (b) some feel for the relationship of variation in test parameters to variation in mission performance. Measurement variations must be significantly less than variations due to other sources, and less than variations which would affect mission performance.

In order to investigate test repeatability and isolate some of the factors which might lead to measurement variation, a small data collection effort was initiated with respect to a selected test point in the AN/ASN-90 depot repair process. The particular test which was selected for this purpose was the final acceptance test. This test was selected because, out of all depot repair and test operations, the final acceptance test represents the single largest cost item.

The data collection design for collecting test repeatability data is described below, followed by a presentation of selected results from implementing the design.

Data Collection Design

Development of the data collection design was influenced by several considerations. First, it was judged important that any design developed provide data to document the extent to which there is, in fact,

a test repeatability problem. Toward this end, any data collection design developed had to yield data that would provide an estimate of the total unreliability of test results, i.e., an estimate of the total variation in repeated measurements for a given component (item) due to the combined influence of all factors in the system which might produce such variation.

Second, it was judged important that any design developed provide information on some of the potential sources of unreliability of test results. These potential sources included: (1) systematic differences between test stations (due to, e.g., calibration differences), (2) systematic differences between operators at test stations, and (3) for a given operator and test station, measurement variation over time for a given component. Systematic differences between test stations and/or operators, if found, would suggest alternative corrective actions to possibly reduce such differences, such as improved operator training, or improved station calibration procedures. On the other hand, variation in test results for a given station and operator could be attributable to several sources, including time to time variations within the component being tested, and/or within the operator, and/or within the test equipment itself. Additional studies would be required to separate out these sources of variation.

A third major consideration in developing the data collection design concerned the number of "extra" tests that would be required, i.e., the number of tests over and above the single test that would normally be made on each component. In turn, the total number of extra tests is a function of the number of components for which repeated measurement data are to be obtained, and the number of repeated measurements on each component. Clearly, any design developed had to hold the number of extra tests within reasonable bounds, in view of the time/costs associated with such tests.

A fourth requirement concerned the number of operators and test stations that would be involved in the data collection, in order to assess any differences between operators and test stations. It was decided that at least three operators and three test stations should be involved.

Several alternative experimental designs were generated and evaluated in view of the above requirements and constraints. One design considered initially was a complete factorial design, with operators, test stations, and components as the three factors, with each component being tested at all operator-station combinations. This design was rejected, however, due to the prohibitively large number of extra tests that would be required, given a sufficiently large sample of components. For example, for nine components, three test stations, and three operators, 81 total tests would be required (even without measurement replication at given station-operator combinations) with 72 of these tests being extra tests.

A second design considered was a Latin Square design, with, for example, three operators, three test stations, and three components, with each component tested at three different station-operator combinations. This type of design was considered undesirable, however, due to the possible confounding of the main effects of stations, operators, and components with interactions involving these factors. Since it was not judged tenable to assume that these interactions were zero, this design was rejected.

A third design considered was a "randomized groups" design, with *n* different components assigned at random to each station-operator combination, and with each component being tested twice at a given station-operator combination. For example, with three operators and three test stations, three different components might be assigned at random to each of the nine station-operator combinations. Although this design is statistically sound, it was judged to be less than optimal since it does not provide for a direct control of differences between components in assessing differences between stations and operators.

The data collection design finally selected (see Figure 1) does provide for control of differences between components in assessing differences between stations and operators, in that the same components are tested at each test station, and the same components tested by each

	Station 1	Station 2	Station 3
Operator 1	C_1 C_2 (2) C_3	C_4 C_5 (2) C_6	C_7 C_8 C_9 (2)
Operator 2	C_7 (2) C_8 C_9	C_1 C_2 C_3 (2)	C_4 C_5 C_6 (2)
Operator 3	C_4 (2) C_5 C_6	C_7 C_8 (2) C_9	C_1 (2) C_2 C_3

FIGURE 1. DATA COLLECTION DESIGN

operator.* The design calls for the testing of a sample of nine components (IMU's) (C_1 through C_9), involving three test stations, and one operator selected within each shift, giving three operators.

The sample of nine components are split at random into three groups of three components each (Group 1--C_1, C_2, C_3; Group 2--C_4, C_5, C_6; Group 3--C_7, C_8, C_9). Each group of components are then tested at three different station-operator combinations, in a way such that all nine components are tested at each station, and such that all nine components are tested by each operator. In addition, each of the nine components are tested twice [indicated by the (2)] at a given station-operator combination, to assess time associated differences in test results.

To implement the design, a total of 36 final acceptance (ATP) tests are required (each of nine components tested four times), or 27 "extra" tests over and above the single tests for each of the nine components normally required.

The test schedule for implementing the design is shown in Figure 2. The schedule shows, for each component (IMU), the station-operator combination at which the first, second, third, and fourth tests occur. The schedule is designed such that the order of testing is balanced with respect to test stations and operators. Also, the two tests for a component at the same operator-station combination are always the third and fourth tests for that component.

In implementing the design, for the first three tests for a given component, only a portion of the ATP was to be completed (approximately the first four hours of testing). Specifically, the first three tests were to begin with the Gyro Slew Checks, and continue through the calculation of Gypto Torquer Scale Factors (see Exhibit A, Data Collection Form). The

---

* This type of design has been labelled as a Type II "mixed design." See Lindquist, E. F., "Design and Analysis of Experiments in Psychology and Education" (Ref 4, pp 273-281). Also see Kirk, R. E., "Experimental Design: Procedures for the Behavioral Sciences" (Ref 2, Chapter 9).

IMU	Serial No.	ORDER Test 1	Test 2	Test 3	Test 4
1		O_1S_1	O_2S_2	O_3S_3	O_3S_3
2		O_2S_2	O_3S_3	O_1S_1	O_1S_1
3		O_3S_3	O_1S_1	O_2S_2	O_2S_2
4		O_1S_2	O_2S_3	O_3S_1	O_3S_1
5		O_2S_3	O_3S_1	O_1S_2	O_1S_2
6		O_3S_1	O_1S_2	O_2S_3	O_2S_3
7		O_1S_3	O_3S_2	O_2S_1	O_2S_1
8		O_2S_1	O_1S_3	O_3S_2	O_3S_2
9		O_3S_2	O_2S_1	O_1S_3	O_1S_3

FIGURE 2. TEST SCHEDULE

AD-A032 205 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OHIO F/G 12/2
PROCEEDINGS OF THE SYMPOSIUM ON APPLICATION OF DECISION THEORY --ETC(U)
OCT 76 N K WOMER

UNCLASSIFIED AFIT-TR-76-15 NL

2 OF 3
AD
A032205

fourth test for each component was to consist of the full ATP, as ordinarily conducted.

The above "partial testing" plan was designed to reduce time/costs associated with the retesting of components. However, it was judged that the partial tests would provide sufficient data for analysis purposes.

The analysis of variance source table for the data collection design is shown as Table 1, which gives a breakdown of sources of variation considering only the first three measurements on each component. The total variation is broken down into two parts: between components, and within components. The between components variation, in turn, breaks down into: (a) variation between groups, and (b) variation between components within groups. For this particular design, the between groups variation reduces to the between component portion of the station by operator interaction, i.e., that part of the operator by station interaction based on between component comparisons [*AB(b)*].*

---

* See Lindquist, (Ref 4).

TABLE 1. ANALYSIS OF VARIANCE TABLE FOR DATA COLLECTION DESIGN

Source of Variation	df	E(MS)
Between Components	$an-1 = 8$	
- $G[AB(b)]$	$a-1 = 2$	$\sigma^2 + a\sigma_C^2 + na\sigma_G^2$
- C within groups	$a(n-1) = 6$	$\sigma^2 + a\sigma_C^2$
Within Components	$an(a-1) = 18$	
- A	$a-1 = 2$	$\sigma^2 + na\sigma_A^2$
- B	$a-1 = 2$	$\sigma^2 + na\sigma_B^2$
- AB(w)	$(a-1)(a-2) = 2$	$\sigma^2 + n\sigma_{AB}^2$
- Residual	$a(a-1)(n-1) = 12$	σ^2
Total	$a^2n-1 = 26$	

A = Test stations

B = Operators

C = Components

G = Groups

a = Number of operators = number of test stations

n = Number of components in each group

The within components variation, which is taken as a measure of the total unreliability of test results, breaks down into: (a) differences between test stations (*A* main effect); (b) differences between operators (*B* main effect); (c) that part of the station by operator interaction based on within components comparisons [*AB*(w)], and (d) residual variation. Part of this residual variation is variation in repeated measurements of a component at a given station-operator combination, which can be calculated based on the third and fourth measurements of a component.

The last column of Table 1 shows the variance components in the expected mean squares. As illustrated subsequently, these expressions can be used to calculate variance estimates for each source of variation.

In summary, repeated measurement data was to be collected and analyzed to provide results that include:

(1) Estimate of the total variation in repeated measurements for a given component due to the combined influence of operator factors, test equipment factors, time associated variation within a component (item) being tested, and other unknown sources of variation

(2) Estimate of changes in "pass-fail" decisions as a result of the variability in (1) above

(3) Estimate of measurement variation over time (for a given operator, station, and component)

(4) Estimate of systematic differences between test stations, indicative of station bias

(5) Estimate of systematic differences between operators, indicative of operator bias

(6) Estimate of differences between components being tested

(7) Estimate of relative variability of test readings between test stations and between operators.

The above estimates were to be obtained for each of the test parameters involved in the ATP for which data are being collected. These results were then be used in the simulation model to investigate potential cost savings from changes.

Test Results

Test data were collected for seven different measures involved in the ATP. Results from one of these measures (z Slew Rate) are presented here. Analogous results for the other six measures are presented in Appendix .

Table 2 shows the analysis of variance results (associated means and standard deviation for each test station and operator are given in Table 3). As indicated in Table 2, the only statistically significant effect was the main effect of stations.* The operator main effect, and the operator by station interaction, both were not statistically significant.

However, of more interest for this problem than statistical significance is the variance estimates for each source of variation. These estimates are shown in Table 4.

For example, the variance estimate for test stations is 13.490. This estimate was calculated using the expressions for the expected mean squares given earlier in Table 1, as follows:

(1) $MS_A = \sigma^2 + na\sigma_A^2$. Substituting in this expression the calculated values for MS_A, σ^2 (residual), and values for the coefficients n and a gives:

(2) $145.45 = 24.04 + (3)(3)\sigma_A^2$

* The error term in the F ratio to test the group effect [$AB(b)$] is the mean square of components within groups. The error term in the F ratio for all the "within" effects is the residual (within) mean square.

TABLE 2. ANALYSIS OF VARIANCE RESULTS FOR Z AXIS SLEW RATE

Source	df	SS	MS	F
Between Components	8	13,563.29	1,695.41	
• *G* [*AB(b)*]	2	1,592.81	796.40	0.399
• *C* within groups	6	11,970.48	1,995.08	
Within Components	18	724.88	40.27	
• *A*	2	290.91	145.45	6.051*
• *B*	2	121.60	60.80	2.529
• *AB* (w)	2	23.91	11.95	0.497
• Residual	12	288.47	24.04	
Total	26	14,288.17		

A = Test Stations

B = Operators

C = Components (IMU's)

G = Groups

* F Ratio significant at the 0.05 level.

TABLE 3. MEANS AND STANDARD DEVIATIONS (Z SLEW RATE) FOR EACH OPERATOR AND TEST STATION

		Mean	Standard Deviation
Station	1	5392.7	23.20
	2	5400.7	24.51
	3	5396.9	24.71
Operator	1	5398.8	21.83
	2	5393.8	25.77
	3	5397.7	25.11

TABLE 4. VARIANCE ESTIMATES FOR Z AXIS SLEW RATE

Source	Variance Estimate	Percent of Total Variance	
Between Components	657.014	94	
• *G* [*AB(b)*]	0.000		
• *C* within *G*	657.014		
Within Components	41.614	6	
• *A*	13.490		2.0
• *B*	4.084		0.6
• *AB*(w)	0.000		0.0
• Residual	24.040		3.4
Total	698.628	100	

A = Test Stations

B = Operators

C = Components (IMU's)

G = Groups

$$(3)\ \sigma_A^2 = \frac{145.45 - 24.04}{9} = \frac{121.41}{9} = 13.490.$$

The other variance estimates are calculated in a similar way.* The sum of each of the six separate variance estimates provides an estimate of the total variance (698.628).

Inspection of these estimates shows that differences between components account by far for the largest portion of the variance - approximately 94 percent of the total variance ($\frac{657.014}{698.628} = 0.9404$). This leaves about 6 percent of the variance attributable to variation in repeated measures onto ($\frac{41.614}{698.628} = 0.596$). That is, approximately 6 percent of the total variance is due to repeated measurements of a given component, when these measurements are made at different station/operator combination.

If the total variation in repeated measurements (41.614), most of this variation can be attributable to factors other than systematic differences between test stations and operators, since the residual variance (24.04) accounts for about 58 percent ($\frac{24.04}{41.614}$) of the variation in repeated measurements. Differences between test stations account for about 32 percent of the variation in repeated measurements, and differences between operators about 10 percent.

Finally, it is of interest to estimate the variance in repeated measurements for a <u>given</u> station/operator combination, i.e., the variation between the third and fourth measurements for a given component, as pointed out earlier. This estimate is calculated to be 15.632, and represents the residual mean square in a variance breakdown with operator and test stations as two factors, with two measurements on each of 9 components assigned at random--to the nine station operator combinations. The estimate of 15.632 can be compared with the residual variance of 24.040 as given in Table 2, indicating that over one-half

---

* Negative variance estimates were obtained for groups (G) and $AB_{(w)}$. Thus, the value of 0.0 was used for each of these two estimates.

of the latter residual variation can be attributable to variation in repeated measurements at a given station/operator combination.

As mentioned earlier, the various estimates such as given above are to serve as imputs to the simulation model, in order to investigate potential cost savings from changes.

REFERENCES

(1) Genet, R. M., Evaluating Inertial Guidance Acceptance and Performance Tests. Aerospace Guidance and Metrology Center, Newark Air Force Station, Ohio, 1970.

(2) Genet, R. M., A Nonlinear Programming Determination of Depot Maintenance Errors. Aerospace Guidance and Metrology Center, Newark Air Force Station, Ohio, 1970.

(3) Kirk, R. E., Experimental Design: Procedures for the Behaviorial Sciences, Brooks/Cole Publishing Company, 1968.

(4) Lindquist, E. F., Design and Analysis of Experiments in Psychology and Education, Houghton Mifflin Company, 1953.

EXHIBIT A. DATA COLLECTION FORM

Start ETI reading ________________ hr

9-27 Gyro slew check

Z slew rate (θZ) Actual reading ________________ deg/hr

X slew rate (θX) Actual reading ________________ deg/hr

Y slew rate (θY) Actual reading ________________ deg/hr

9-31 Azimuth grid mode slew

Kwe sinλ Actual reading ____________ deg/hr/V

9-35 Multi position tests

GCM Position 1 biases

X														
Y														
Z														

9-55 X Gypto Torquer Scale Factor (KTX)	+	.						arc sec/pulse
Y Gypto Torquer Scale Factor (KTY)	+	.						arc sec/pulse
Z Gypto Torquer Scale Factor (KTZ)	+	.						arc sec/pulse

End ETI reading ________________ hr

FAILURE DETECTION AIDS FOR HUMAN OPERATOR DECISIONS IN A PRECISION INERTIAL NAVIGATION SYSTEM COMPLEX

Thomas H. Kerr

Member of the Technical Staff
The Analytic Sciences Corporation
Reading, Massachusetts

An existing redundant standby system consisting of two identical Inertial Navigation Systems (INS) may be improved by using a third INS of higher precision as a monitor. The monitor INS provides internal reset corrections which allow the time between external reference position fixes to be extended. Conventional failure detection aids have evolved to enable a human operator to select the best standby INS for navigation. A new method for detecting failures in the monitoring INS has been developed and its performance has been evaluated. A 54 discrete state Markov probability model of the total three-INS complex as a function of the respective Mean-Time-To-Failure and Mean-Repair-Times is presented in this paper. A technique is included for identifying and modeling the human operator's decisions as non-ideal switches, with/without the new method of failure detection for the monitoring INS. This Markov probability model is used as an intermediate step in explicitly specifying how the new failure detection method should interface with the total three-INS complex. This is accomplished by quantitatively evaluating the total system availability. Then the policy which results in the greatest system availability is selected as the most desirable.

1. INTRODUCTION

A simplified block diagram of a navigation system, consisting of three Inertial Navigation Systems (INS), is portrayed in Fig. 1. This configuration represents an improvement to the previously developed standby redundant INS1/INS1' navigation system through the use of the higher precision INS2 as a monitor to enable internal reset corrections between external position fixes. Although the INS contains precision gyros, they incur sufficient drift to necessitate a periodic corrective action consisting of an external position fix. Use of the higher precision INS2 for internal corrections allows a longer time between external fixes to maintain the same or better total system navigation accuracy. The monitoring INS2 is assumed to only have the higher accuracy needed to supply corrective resets to INS1 gyro drift-rates and <u>not</u> to have full navigation capabilities. A more detailed view of the operation and sampling times of this three INS complex is afforded in Fig. 2 where both the Inertial Navigation Unit and its associated controlling software consisting of Kalman filters based on linear error models of the component gyros are shown. (The use of Kalman filters in this role is standard as discussed in Ref. 1.)

An operator (navigator) oversees the operation of this three INS complex. This human operator makes the following decisions which are respectively represented by switches S' and S of Figs. 1 and 2:

- operator selects the mode of operation (either prime mode or back-up mode)
- operator selects the master on-line INS1.

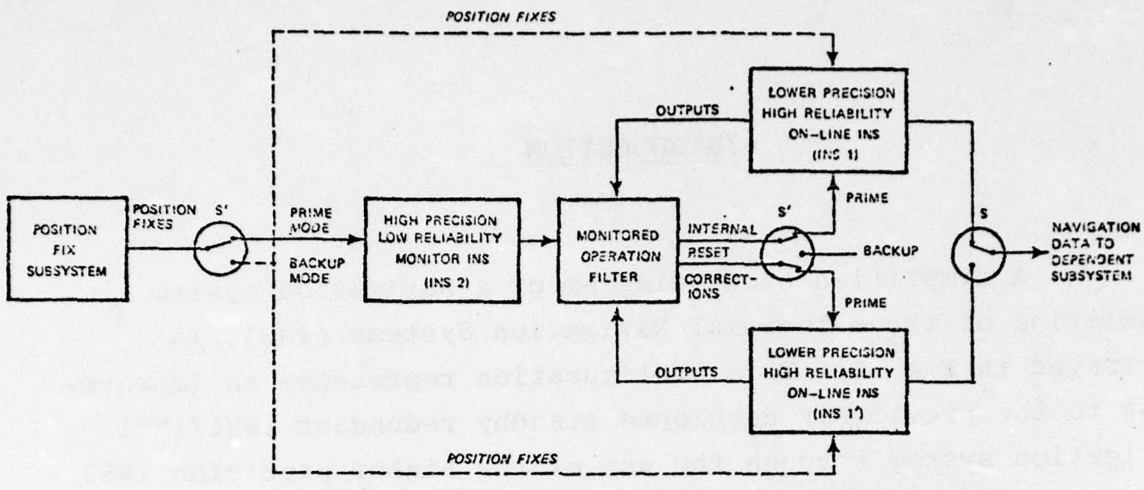


Figure 1 Simplified Block Diagram Overview of the Three INS Complex

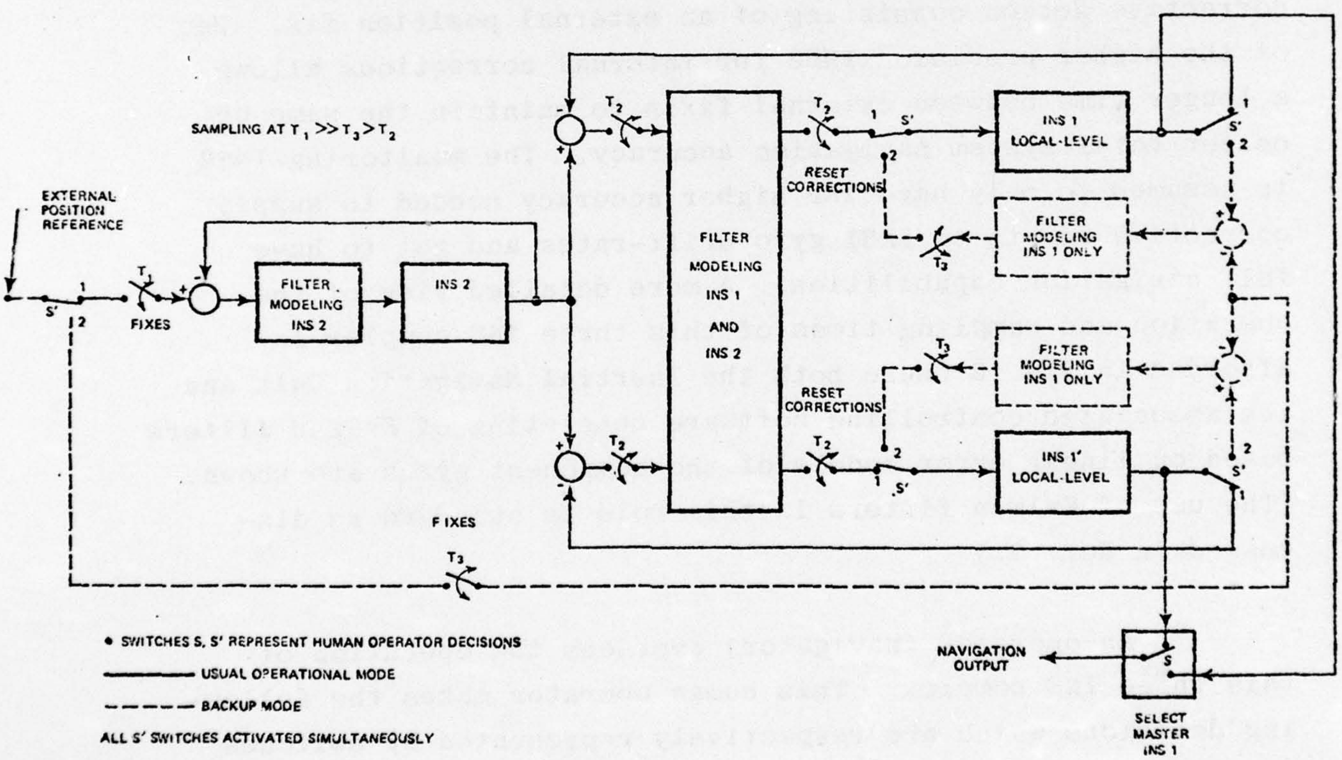


Figure 2 More Detailed View of the INS1/INS2 System

The prime mode uses the newly configured INS1/INS2 while the back-up mode (represented by the dashed lines in Figs. 1 and 2) uses only the standard INS1/INS1'. However, both modes require the selection of the master on-line INS1. The operator's selection of the master INS1 is based on the following information on each INS1 as obtained from a computer or technician's graph:

- plots of the divergence of INS1 (INS1') from external fixes
- plots of INS1 (INS1') gyro bias correction histories
- plots of INS1 (INS1') comparisons to any available reference.

After the total problem of implementation had been partitioned into tractable sub-problems, our primary objective was to specify a method for detecting failures in the INS2. It is critical to detect any INS2 failures, otherwise, faulty internal reset corrections are provided to both INS1 and INS1'. Once the method for detecting failures in the INS2 had been specified and the performance had been evaluated, a natural consequence of the completion of the primary objective is encountered as the secondary objective of specifying how the INS2 failure detector will interface with the entire system. The interface should also include any additional INS2 failure detection aids that are subsequently developed to help the navigator decide in what mode he should operate the system.

The underlying framework of an analysis of the three INS system of Figs. 1 and 2 is presented herein in terms of the simplified representation of Fig. 3, using the standard reliability theory techniques of Ref. 2. The use of more realistic non-ideal switches to represent the human operator's function gives rise to 54 discrete states of a Markov

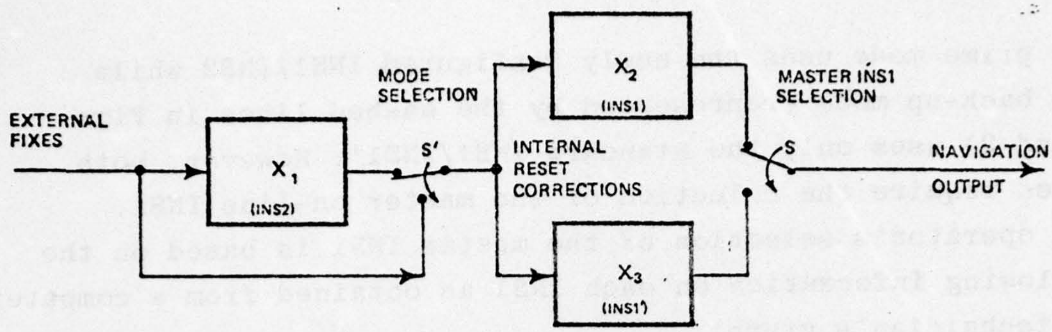


Figure 3 Block Diagram of Three INS Complex as Used in the Availability Analysis

probability model of the system as is discussed in Section 3. A technique is proposed for relatively weighting the failure monitoring test results of the specified INS2 failure detection method with those of any other INS2 failure detection aid (manual or automatic) that is subsequently developed. A quantification of the total system availability (i.e., reliability with repair) is then obtained and the best interfacing policy for integrating the proposed INS2 failure detector into the total system is the policy (corresponding to a weighting) that yields the greatest system availability.

2. AN INS2 FAILURE DETECTION METHOD

A summarizing overview is given in Fig. 4 of how this failure detection approach works. The theoretical basis of the so-called CR2* failure detection approach is a generalization of the use of confidence intervals. The three main ideas that serve as the foundation for CR2 failure detection are shown as three different diagrams in Fig. 4 and are discussed below. These diagrams are shown in juxtaposition to facilitate a comparison of how the relative overlapping of the confidence regions affects the scalar test statistic at three specific check times, t_1, t_2, and t_3. These confidence regions are portrayed in the top diagram of Fig. 4. At each check time, these confidence regions are elliptical. <u>A failure is declared when the two confidence regions do not overlap</u>.

At each check time, t_i, the two elliptical cross sections of the confidence regions, shown in the top diagram of Fig. 4, are fixed levels of two paraboloids, shown in the middle diagram. The problem is to determine whether these two ellipses overlap. In developing the real-time detection algorithm, the <u>test for the presence or absence of overlap was formulated as the solution of a minimization problem.</u> The relative position of $\ell(t_i)$, the minimum point of the intersection of the two paraboloids, to $K_1(t_i)$, the level that corresponds to the elliptical cross section of the confidence regions, determines if there is overlap and,

*CR2 is an acronym for Two (2) Confidence Regions.

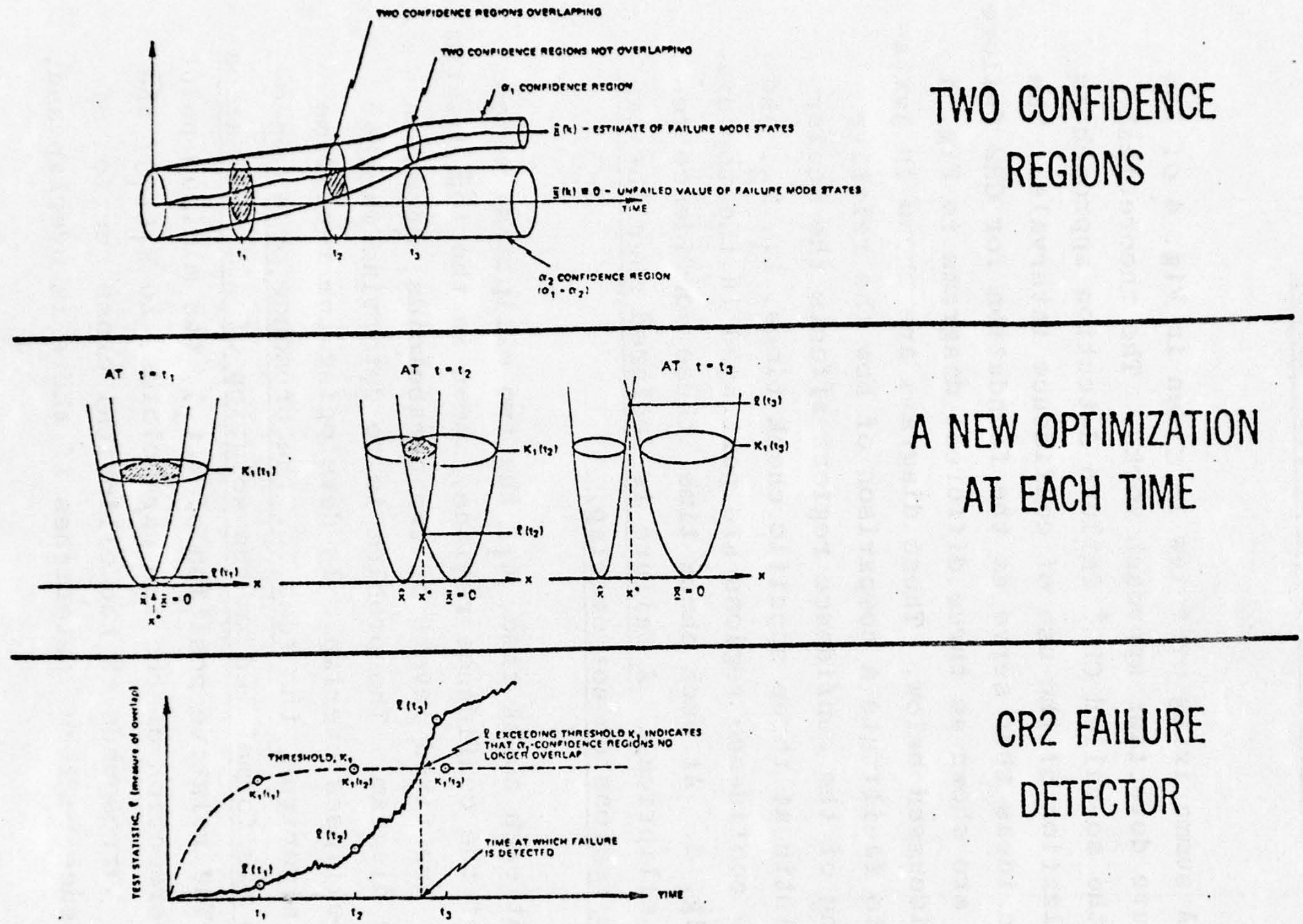


Figure 4 Overview Shows How Comparison of Scalar Test Statistic to Decision Threshold Relates to Test for Confidence Region Overlap Through Artifice of Optimization Problem

if so, the amount of overlap. As long as $\ell(t_i)$ is below $K_1(t_i)$ there is overlap, but when $\ell(t_i)$ exceeds $K_1(t_i)$, <u>then the confidence regions are disjoint and a failure is declared.</u> The relationship between the test statistic $\ell(t_i)$ and the decision threshold $K_1(t_i)$ is summarized in the bottom diagram of Fig. 4. <u>It is sufficient to observe only the test statistic $\ell(t_i)$, and to declare failures when $\ell(t_i)$ exceeds $K_1(t_i)$.</u>

In the CR2 failure detector, a higher level of the threshold K_1, to which the test statistic $\ell(t_i)$ is compared, effectively raises the heights of the ellipses in the associated optimization problem; this corresponds to stouter confidence regions. Analytic expressions have been derived which are used for pre-specifying the time-varying decision threshold K_1 and for evaluating the instantaneous probabilities of the test statistic exceeding the threshold under H_0 (no-failure) and H_1 (a particular magnitude failure), respectively, as P_{fa} and P_d. The expressions are used in the setting of the threshold K_1 in a characteristic trade-off of instantaneous probability of false alarm P_{fa} versus the probability of correct detection P_d* associated with hypothesis testing detection decisions. Early derivations of the CR2 failure detection approach may be found in Refs. 3 and 4, while the most recent theoretical refinements and convergence and convergence-rate proofs are in Ref. 5. The mathematical techniques of confidence regions used in these derivations are similar to those used in Ref. 6.

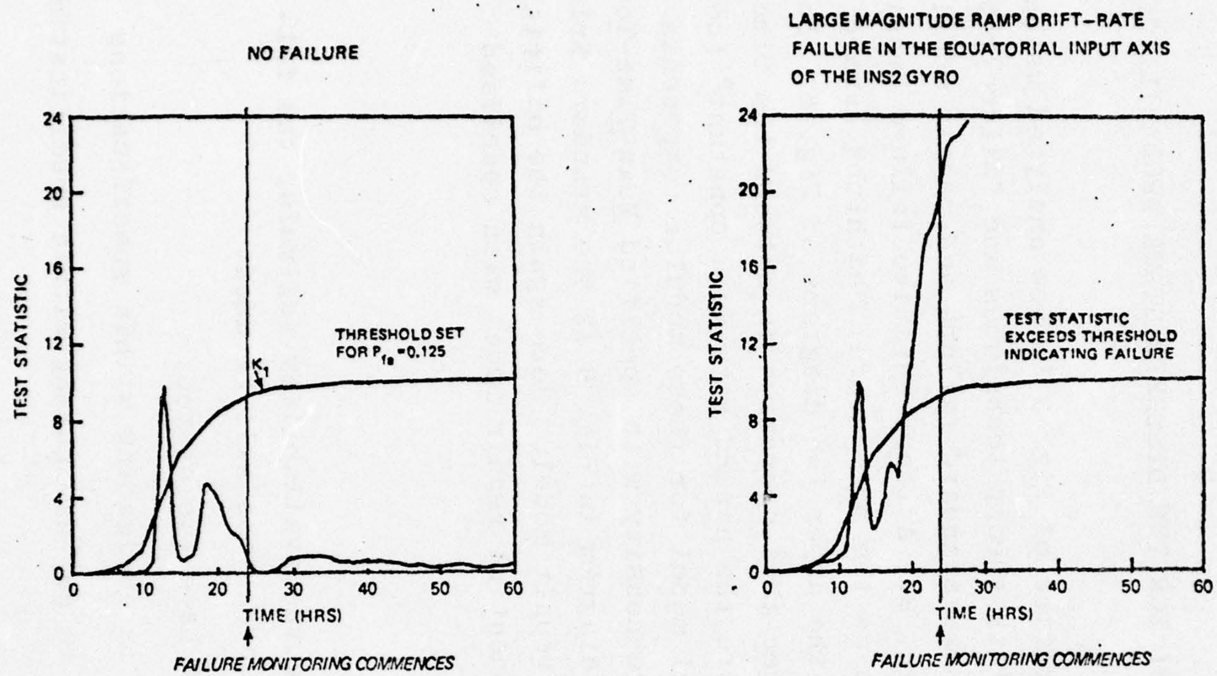
*The probability of miss is $P_m=1-P_d$.

Unlike the situations where the usual likelihood ratio is a Uniformly Most Powerful (UMP) test in that it is as good as or better than any other decision test, an INS2 failure represents a random event that occurs at a random or unknown time. There is little special justification for using a likelihood ratio as a decision function (Ref. 7, p. 96; Ref. 8, p. 315) since it is not UMP for this situation and there may be other decision functions that are as good or better. Confidence region tests serve as one alternative. Recently, confidence region approaches have been developed for other detection applications as well (Ref. 9).

When external position fixes are available to the three-INS complex, failures in the individual INS are more easily detected by a comparison to this more accurate external fix as a standard. The CR2 failure detector is used to detect failures (of a certain critical magnitude corresponding to a certain critical Signal-to-Noise Ratio associated with the resulting failure signal response) between external fixes.

The CR2 failure detection performance on real INS2 data is presented in Fig. 5. The plot on the left of Fig. 5 represents the CR2 test statistic and the pre-specified decision threshold K_1 under a no-failure condition. After an initial 24 hrs,* the test statistic is well below the decision threshold, confirming that no failure is present. The plot on the right of Fig. 5 represents the CR2 test statistic and decision threshold for a large magnitude ramp drift-rate failure in one of the INS2 gyros. Following the initial 24 hr waiting period, the CR2 test statistic quickly exceeds the threshold correctly indicating a failure.

*The error dynamics of a gyro reflect the 24 hr earth rotation rate; consequently, approximately one full 24 hr period is required for the filter to lock onto the cycle of the underlying sinusoid.



- SIMULATED FAILURE AT t = 0 HRS
- IGNORING INITIAL TRANSIENTS DURING FIRST 24 HRS FOLLOWS POLICY SPECIFICATION

Figure 5 INS2 (CR2) Failure Detection Performance Using Real Data

3. AVAILABILITY ANALYSIS OF THE THREE INS COMPLEX

3.1 THE MARKOV MODEL FOR THE DISCRETE STATE PROBABILITIES

The configuration of Fig. 3 is now analyzed using the standard reliability theory techniques and failure/repair models of Ref. 2. These standard failure and repair models are illustrated in Fig. 6. A representative failure density and its normalization as the more useful "bathtub" hazard rate are depicted in the upper two diagrams of Fig. 6. For each INS, it is assumed that operation is between the times t_1 and t_2 and that here the hazard rate is a constant* (consistent with the usual model for these modular components as an exponential failure density with specified Mean-Time-To-Failure). The lower diagram in Fig. 6 is the standard Erlang or gamma distributed repair model. Here again the official Navy model is an exponential repair model with specified Mean-Repair-Time.

In performing the availability analysis, the following notation and conventions will be used.

Modular INS components have two states:

$X_i \triangleq$ component i is navigating within specifications

$\overline{X}_i \triangleq$ component i is not navigating within specifications.

---

*A constant hazard rate is reasonable assuming that infant mortality failures have been weeded out through initial acceptance testing and that wear-out failures are avoided through periodic preventive maintenance.

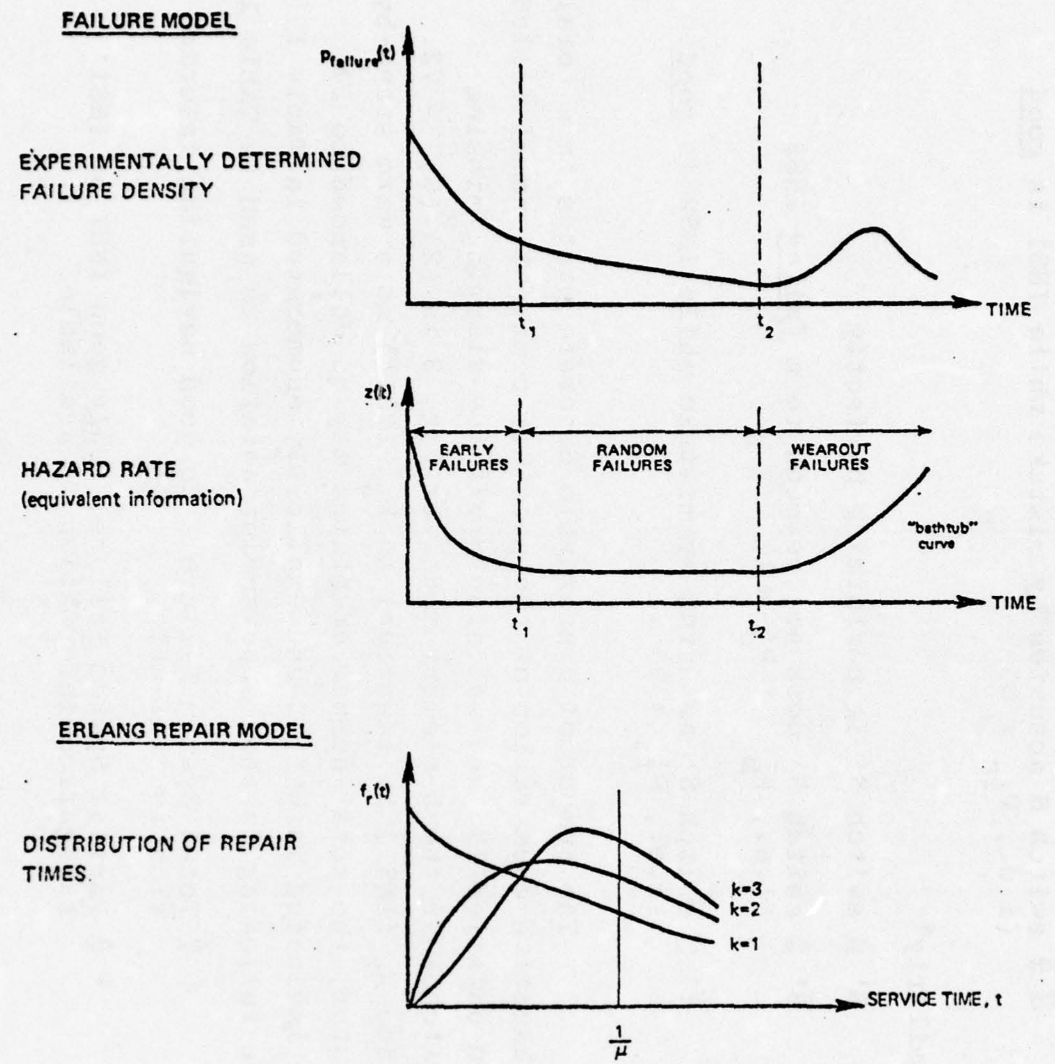


Figure 6 Standard Failure and Repair Models Used in the Availability Analysis

The non-ideal switches representing the human operator decisions have three states:

$S \triangleq$ switch S is performing perfectly

$\bar{S} \triangleq$ switch S does not respond to a failed INS1 (i.e., $P_M \equiv 1-P_d < 1$)

$\bar{\bar{S}} \triangleq$ switch S actuates by mistake while INS1 is good (i.e., $P_{fa} > 0$) .

Similarly*

$S' \triangleq$ switch S' is performing perfectly

$\bar{S}' \triangleq$ switch S' does not respond to a failed INS2 (i.e., $P'_M \equiv 1-P'_d < 1$)

$\bar{\bar{S}}' \triangleq$ switch S' actuates by mistake while INS2 is good (i.e., $P'_{fa} > 0$) .

The number of discernable discrete states in a totally exhaustive enumeration of the possible combinations of failed and unfailed INS and satisfactory/false-alarming/missing switches in the 5-element model of Fig. 3 is $(2)^3(3)^2 = 72$. Since X_2 (INS1) is identical to X_3 (INS1') as a warm stand-by system, the total number of states may be collapsed to 54 as indicated in Ref. 2 and explicitly enumerated in Table 1. The following further superscript notation is used in Table 1:

† $\triangleq$ Total System Failure. No good navigation information is available.

* $\triangleq$ Partial System Failure. Only good INS1 or INS1' navigation information is available.

*This reliability analysis is applicable to any INS2 failure detection method that uses the same filter inputs as used by the CR2 test as long as the P'_d and P'_{fa} which characterize it have been evaluated.

TABLE 1

AN EXHAUSTIVE ENUMERATION OF THE DISCRETE STATES IN THE STANDBY/SERIES 5 ELEMENT SYSTEM OF FIG. 3

$S_1 = S'X_1X_2X_3S$	$S_{19} = S'X_1X_2X_3\bar{S}$	$S_{37} = \bar{\bar{S}}'X_1X_2X_3\bar{S}^*$
$S_2 = S'\bar{X}_1X_2X_3S^*$	$S_{20} = S'\bar{X}_1X_2X_3\bar{S}^*$	$S_{38} = \bar{\bar{S}}'\bar{X}_1X_2X_3\bar{S}$
$S_3 = S'X_1\bar{X}_2X_3S + S'X_1X_2\bar{X}_3S$	$S_{21} = S'X_1\bar{X}_2X_3\bar{S} + S'X_1X_2\bar{X}_3\bar{S}^\dagger$	$S_{39} = \bar{\bar{S}}'X_1\bar{X}_2X_3\bar{S} + \bar{\bar{S}}'X_1X_2\bar{X}_3\bar{S}^\dagger$
$S_4 = S'\bar{X}_1\bar{X}_2X_3S + S'\bar{X}_1X_2\bar{X}_3S^*$	$S_{22} = S'\bar{X}_1\bar{X}_2X_3\bar{S} + S'\bar{X}_1X_2\bar{X}_3\bar{S}^\dagger$	$S_{40} = \bar{\bar{S}}'\bar{X}_1\bar{X}_2X_3\bar{S} + \bar{\bar{S}}'\bar{X}_1X_2\bar{X}_3\bar{S}^\dagger$
$S_5 = S'X_1\bar{X}_2\bar{X}_3S^\dagger$	$S_{23} = S'X_1\bar{X}_2\bar{X}_3\bar{S}^\dagger$	$S_{41} = \bar{\bar{S}}'X_1\bar{X}_2\bar{X}_3\bar{S}^\dagger$
$S_6 = S'\bar{X}_1\bar{X}_2\bar{X}_3S^\dagger$	$S_{24} = S'\bar{X}_1\bar{X}_2\bar{X}_3\bar{S}^\dagger$	$S_{42} = \bar{\bar{S}}'\bar{X}_1\bar{X}_2\bar{X}_3\bar{S}^\dagger$
$S_7 = \bar{S}'X_1X_2X_3S$	$S_{25} = S'X_1X_2X_3\bar{\bar{S}}$	$S_{43} = \bar{S}'X_1X_2X_3\bar{\bar{S}}$
$S_8 = \bar{S}'\bar{X}_1X_2X_3S^\dagger$	$S_{26} = S'\bar{X}_1X_2X_3\bar{\bar{S}}^*$	$S_{44} = \bar{S}'\bar{X}_1X_2X_3\bar{\bar{S}}^\dagger$
$S_9 = \bar{S}'X_1\bar{X}_2X_3S + \bar{S}'X_1X_2\bar{X}_3S$	$S_{27} = S'X_1\bar{X}_2X_3\bar{\bar{S}} + S'X_1X_2\bar{X}_3\bar{\bar{S}}^\dagger$	$S_{45} = \bar{S}'X_1\bar{X}_2X_3\bar{\bar{S}} + \bar{S}'X_1X_2\bar{X}_3\bar{\bar{S}}^\dagger$
$S_{10} = \bar{S}'\bar{X}_1\bar{X}_2X_3S + \bar{S}'\bar{X}_1X_2\bar{X}_3S^\dagger$	$S_{28} = S'\bar{X}_1\bar{X}_2X_3\bar{\bar{S}} + S'\bar{X}_1X_2\bar{X}_3\bar{\bar{S}}^\dagger$	$S_{46} = \bar{S}'\bar{X}_1\bar{X}_2X_3\bar{\bar{S}} + \bar{S}'\bar{X}_1X_2\bar{X}_3\bar{\bar{S}}^\dagger$
$S_{11} = \bar{S}'X_1\bar{X}_2\bar{X}_3S^\dagger$	$S_{29} = S'X_1\bar{X}_2\bar{X}_3\bar{\bar{S}}^\dagger$	$S_{47} = \bar{S}'X_1\bar{X}_2\bar{X}_3\bar{\bar{S}}^\dagger$
$S_{12} = \bar{S}'\bar{X}_1\bar{X}_2\bar{X}_3S^\dagger$	$S_{30} = S'\bar{X}_1\bar{X}_2\bar{X}_3\bar{\bar{S}}^\dagger$	$S_{48} = \bar{S}'\bar{X}_1\bar{X}_2\bar{X}_3\bar{\bar{S}}^\dagger$
$S_{13} = \bar{\bar{S}}'X_1X_2X_3S^*$	$S_{31} = \bar{S}'X_1X_2X_3\bar{S}$	$S_{49} = \bar{\bar{S}}'X_1X_2X_3\bar{\bar{S}}^*$
$S_{14} = \bar{\bar{S}}'\bar{X}_1X_2X_3S^*$	$S_{32} = \bar{S}'\bar{X}_1X_2X_3\bar{S}^\dagger$	$S_{50} = \bar{\bar{S}}'\bar{X}_1X_2X_3\bar{\bar{S}}^*$
$S_{15} = \bar{\bar{S}}'X_1\bar{X}_2X_3S + \bar{\bar{S}}'X_1X_2\bar{X}_3S^*$	$S_{33} = \bar{S}'X_1\bar{X}_2X_3\bar{S} + \bar{S}'X_1X_2\bar{X}_3\bar{S}^\dagger$	$S_{51} = \bar{\bar{S}}'X_1\bar{X}_2X_3\bar{\bar{S}} + \bar{\bar{S}}'X_1X_2\bar{X}_3\bar{\bar{S}}^\dagger$
$S_{16} = \bar{\bar{S}}'\bar{X}_1\bar{X}_2X_3S + \bar{\bar{S}}'\bar{X}_1X_2\bar{X}_3S^*$	$S_{34} = \bar{S}'\bar{X}_1\bar{X}_2X_3\bar{S} + \bar{S}'\bar{X}_1X_2\bar{X}_3\bar{S}^\dagger$	$S_{52} = \bar{\bar{S}}'\bar{X}_1\bar{X}_2X_3\bar{\bar{S}} + \bar{\bar{S}}'\bar{X}_1X_2\bar{X}_3\bar{\bar{S}}^\dagger$
$S_{17} = \bar{\bar{S}}'X_1\bar{X}_2\bar{X}_3S^\dagger$	$S_{35} = \bar{S}'X_1\bar{X}_2\bar{X}_3\bar{S}^\dagger$	$S_{53} = \bar{\bar{S}}'X_1\bar{X}_2\bar{X}_3\bar{\bar{S}}^\dagger$
$S_{18} = \bar{\bar{S}}'\bar{X}_1\bar{X}_2\bar{X}_3S^\dagger$	$S_{36} = \bar{S}'\bar{X}_1\bar{X}_2\bar{X}_3\bar{S}^\dagger$	$S_{54} = \bar{\bar{S}}'\bar{X}_1\bar{X}_2\bar{X}_3\bar{\bar{S}}^\dagger$

* = Partial System Failure

† = Total System Failure

It may be summarized from Table 1 that there are

- 9 states for good INS1/INS2 navigation information
- 11 states for good INS1 (or INS1') navigation information only
- 34 states for having no* good navigation information

for a total of 54 discrete states.

*It is considered that no good navigation information is available when both INS1 and INS1' are failed even though the INS2 may be unfailed because the INS2 is assumed to only have higher precision gyros and not to have full navigation capabilities. Therefore, as indicated in Fig. 3, it may not be used alone for navigation.

Associated with this 5-element system is a vector of probabilities corresponding to the likelihood that the system of Fig. 3 will be found in a particular state of the 54 discrete states enumerated in Table 1 at a specified time. In general, the probabilities of occupying a particular state change as a function of time and the specific manner of interconnection and satisfy a linear difference equation of the following form:

$$\begin{bmatrix} P_{S_1}(t+\Delta) \\ P_{S_2}(t+\Delta) \\ \cdot \\ \cdot \\ \cdot \\ P_{S_{54}}(t+\Delta) \end{bmatrix} = \begin{bmatrix} \\ \\ \text{TRANSITION} \\ \text{PROBABILITIES} \\ \\ \end{bmatrix} \begin{bmatrix} P_{S_1}(t) \\ P_{S_2}(t) \\ \cdot \\ \cdot \\ \cdot \\ P_{S_{54}}(t) \end{bmatrix} \qquad (1)$$

This equation characterizes the time evolution of the absolute probability of each of the 54 possible states being assumed by the system as determined by the transition probabilities. An exhaustive enumeration of the 54 states in Table 1 which the system of Fig. 3 can assume is needed as a necessary intermediate step in specifying the elements of the associated transition probability matrix, T in Eq. (1), that characterizes the specific system interconnections and manner of use.

Two representative examples are presented here to demonstrate how the elements of the transition probability matrix are determined. Consider the following state as the focus of attention

$$S_{16} = \bar{\bar{S}}'\bar{X}_1\bar{X}_2X_3S+\bar{\bar{S}}'\bar{X}_1X_2\bar{X}_3S \qquad (2)$$

which represents switch S' in a mode where it would be prone to false alarm (whether X_1 is failed or not), but with X_1 (INS2) in a condition of failure, and either X_2(INS1) or X_3(INS1') in a failed condition (but not both). The state S_{16} is one designated as a partial systems failure because there is one INS1 available and switch S is neither in the false alarm mode nor in the miss mode so the selection of the proper INS1 (INS1') to use in navigation is correctly made and good navigation information is available but only from an INS1 (INS1').

In the flow diagram of Fig. 7,* the state S_{16} may be entered at time = t+Δ (notation: $S_{16}(t+\Delta)$) only through the paths shown from the five states S_4, S_{14}, S_{15}, S_{16}, S_{18} if they were occupied at time = t† (notation: $S_4(t)$, $S_{14}(t)$, etc.). The transition probability, t_{ji}, of going from one state S_i to a distinctly different state S_j is the feed-forward hazard rate times the time step Δ, e.g., the transition probability of going from state $S_{14}(t)$ to $S_{16}(t+\Delta)$ is

$$t_{16,14} = 2\lambda_1 \cdot \Delta \qquad (3)$$

The probability of occupying the same state at time = t+Δ, $S_i(t+\Delta)$, as was occupied on the previous step at time = t, $S_i(t)$, is‡

*Reciprocals of MTTF and Mean-Repair-Time for the INS1 and INS2 are, respectively, λ_1, μ_1 and λ_2, μ_2.

†This model uses the standard reliability assumption of only one allowable change in status during one transition step=Δ.

‡These are elements along the principal diagonal of the transition probability matrix, T.

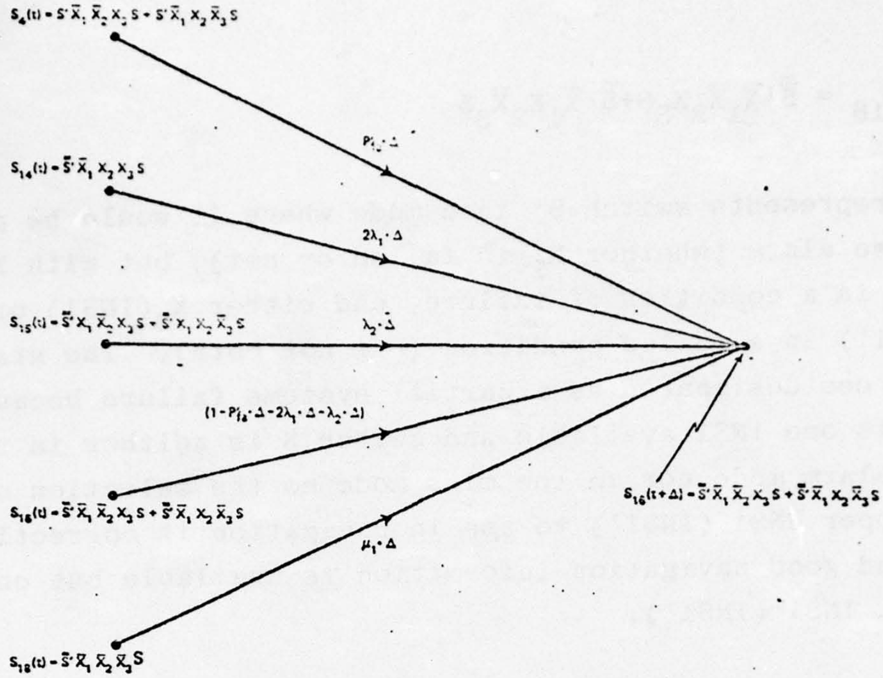


Figure 7 Allowable States at Time=t That May Enter State S_{16} at Time=t+Δ

$$t_{i,i} = 1 - \sum (j^{th} \text{ hazard rate}) \cdot \Delta \tag{4}$$

j ranges over all feedforward paths for switches and failures connecting with state S_i, $j \neq i$

e.g., the transition probability of being in state S_{16} at time = t+Δ given that state S_{16} was occupied at time = t is

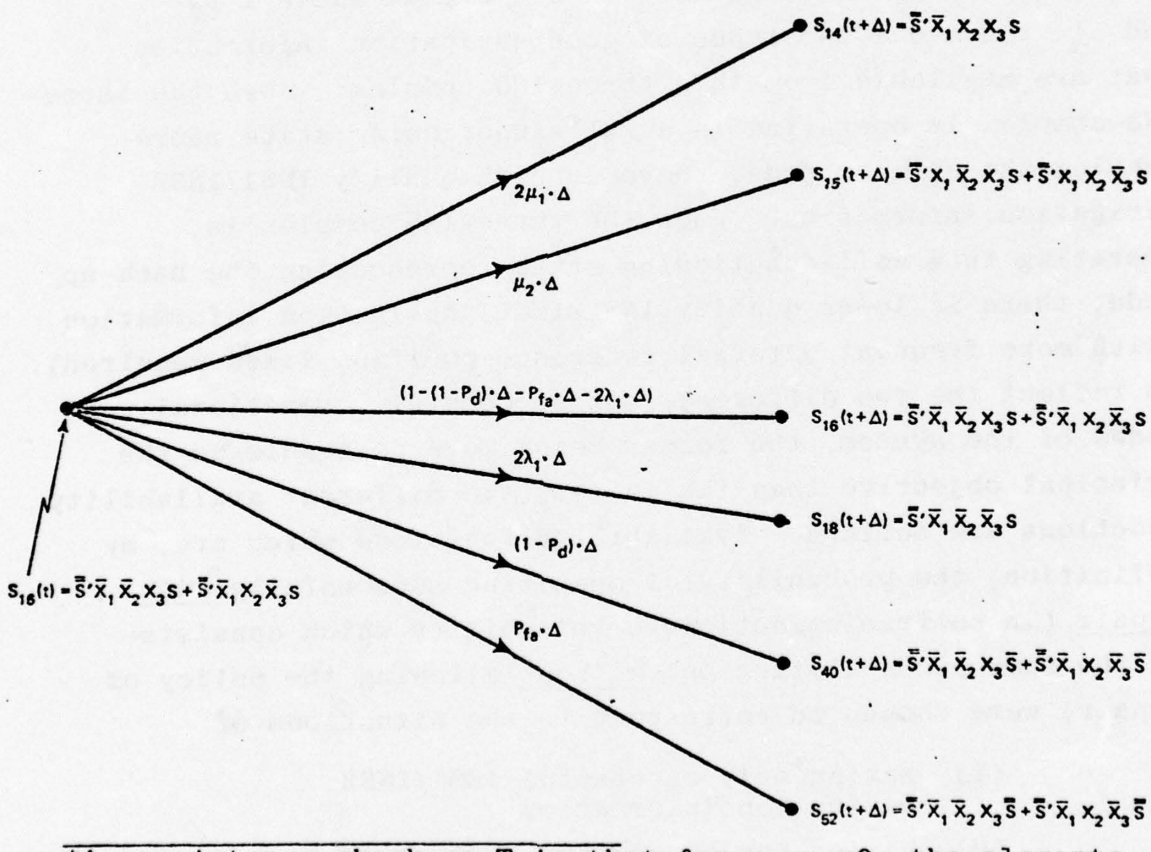
$$t_{16,16} = 1 - P'_{fa} \cdot \Delta - 2\lambda_1 \cdot \Delta - \lambda_2 \cdot \Delta \tag{5}$$

The scalar iteration equation for the probability of being in state S_{16} at time = t+Δ, as obtained from Fig. 7, is

$$P_{S_{16}}(t+\Delta) = P'_{fa} \cdot \Delta \, P_{S_4}(t) + 2\lambda_1 \cdot \Delta \, P_{S_{14}}(t) + \lambda_2 \cdot \Delta \, P_{S_{15}}(t)$$
$$+ (1 - P'_{fa} \cdot \Delta - 2\lambda_1 \cdot \Delta - \lambda_2 \cdot \Delta) \, P_{S_{16}}(t) + \mu_1 \cdot \Delta \, P_{S_{18}}(t) \tag{6}$$

From Fig. 8, the state S_{16} at time = t is an allowable entry state for states $S_{14}, S_{15}, S_{16}, S_{18}, S_{40}, S_{52}$ at time = t+Δ. Therefore, the column of the transition probability matrix T, to be multiplied by $P_{S_{16}}(t)$ in Eq. (1), has non-zero path entries only on the rows corresponding respectively to $S_{14}, S_{15}, S_{16}, S_{18}, S_{40}, S_{52}$. This same type of analysis is performed for each of the other 54 states in order to completely specify the transition probability matrix T.*

As indicated by the above representative example of state S_{16}, T is a rather sparse matrix. Also, notice that the elements of T are the constants $P_d, P_{fa}, P'_d, P'_{fa}, \lambda_1, \lambda_2, \mu_1, \mu_2$; this result is a consequence of the exponential repair models



*A consistency check on T is that for $\mu_1 = \mu_2 = 0$, the elements in each column sum to one.

and the constant hazard functions of the exponential failure models. Since the transition probabilities are all independent of time and depend only on Δ, the underlying Markov process is termed homogeneous. Since the homogeneous process of this analysis does not have the property that any state may be reached directly from any other state with positive probability at each time, this system is not ergodic (Ref. 2). The vector iteration equation of Eq. (1) is linear and time invariant for the system of this analysis and is therefore extremely convenient to work with.

3.2 FORMULATING TWO APPROPRIATE AVAILABILITY EXPRESSIONS

As indicated in Section 3.1 and in Table 1 by * and † there are two types of good navigation information that are available from this three-INS complex. When the three-INS complex is operating in a well-functioning state representing the primary mode, there is high quality INS1/INS2 navigation information. When the three-INS complex is operating in a well-functioning state representing the back-up mode, there is lower quality INS1/INS1' navigation information (with more frequent external reference position fixes required). To reflect the two different, but successful, operational modes of the system, the former being more desirable as the principal objective than the latter, two different availability functions are defined. Availability functions which are, by definition, the probability of operating successfully with repair (in contradistinction to reliability which consists of the same probabilities but without allowing the policy of repair) were chosen to correspond to the situations of

(1) having only successful INS1/INS2 navigation information

(2) having any successful navigation information, either INS1/INS2 or just INS1/INS1'.

These two availability expressions corresponding, respectively, to 1 and 2 above, are

$$A_{1/2}(t) \triangleq P_{S_1}(t) + P_{S_3}(t) + P_{S_7}(t) + P_{S_9}(t) + P_{S_{19}}(t) + P_{S_{25}}(t) + P_{S_{31}}(t) + P_{S_{38}}(t) + P_{S_{43}}(t) \quad (7)$$

$$A_{1/1'}(t) \triangleq A_{1/2}(t) + P_{S_2}(t) + P_{S_4}(t) + P_{S_{13}}(t) + P_{S_{14}}(t) + P_{S_{15}}(t) + P_{S_{16}}(t) + P_{S_{20}}(t) + P_{S_{26}}(t) + P_{S_{37}}(t) + P_{S_{49}}(t) + P_{S_{50}}(t) \quad (8)$$

where the $P_{S_i}(t)$ are the unconditional probabilities of being in the states S_i at time t, with the evolution being completely specified by the iteration equation of Eq. (1). Notice that $A_{1/1}'$, the availability of any good navigation information whatsoever, wholly contains $A_{1/2}(t)$, the availability of the INS1/INS2, in addition to the non-negative quantities representing the unconditional probabilities of being in the states designated as partial failures.

3.3 A PROPOSED NEW MODEL FOR THE HUMAN OPERATOR DECISION

It is standard practice to model the decisions of a human operator as a switch (Ref. 2). When a more detailed and accurate analysis is warranted, it is also standard to acknowledge that the switch is not really perfect (Ref. 10) but inherently contains a detector which may sometimes miss (i.e., switch remains in position because of a failure to recognize that conditions are such that it should have been

thrown) and false alarm (i.e., switch has flipped even though conditions do not warrant it). The ensemble statistics of the imperfect switch which completely characterize its behavior are its inherent probability of false alarm and its probability of miss.

Switch S represents the human operator's decision in the selection of the master INS1, as aided by the information from the accessible computer outputs and plots mentioned in Section 1. Switch S' represents the human operator's decision in the selection of the operational mode. To date, the only INS2 failure detection technique available for use as the detection element (between external position resets) is of the form of the CR2 technique and consequently, its P_d and P_{fa} characteristics apply directly to switch S' only if it is used exclusively. For the case where the CR2 technique is used exclusively, the effective characterizing ensemble statistics for switch S' are:

$$P'_{fa},\text{eff} = P^{OP}_{fa},\text{CR2} \qquad (9)$$

$$P'_{d},\text{eff} = P^{OP}_{d},\text{CR2} \qquad (10)$$

where P'_{fa},eff (P'_d,eff) represents the <u>eff</u>ective P_{fa}(P_d) of switch S' and P_{fa},CR2 (P_d,CR2) represents the particular characterizing statistic of the CR2 failure detector with the <u>O</u>perating <u>P</u>oint (OP) being determined by the level of the pre-specified decision threshold K_1 that is used. However, other INS2 failure detection techniques (possibly manual) are currently under development and upon completion may also be used in the detection element of switch S'. If these new techniques are used exclusively, then the effective characterizing ensemble statistics for switch S' are

$$P'_{fa},eff = P^{OP'}_{fa},manual \tag{11}$$

$$P'_{d},eff = P^{OP'}_{d},manual \tag{12}$$

where the subscript "manual" is just used to distinguish these characteristics from those associated with the CR2 technique.

A natural question to ask is "can the two detection techniques be used together to enhance performance and if so, what would be an optimum mix?" Any attempt to answer this question requires a model for the effective P_d and P_{fa} of switch S' when both failure detection methods are used. The new model that is being proposed to effectively characterize switch S' when both failure detection techniques are being used simultaneously with relative weightings $(1-\lambda)$ and λ is

$$P'_{fa},eff = (1-\lambda)\cdot P^{OP'}_{fa},manual + \lambda\cdot P^{OP}_{fa},CR2 \tag{13}$$

$$P'_{d},eff = (1-\lambda)\cdot P^{OP'}_{d},manual + \lambda\cdot P^{OP}_{d},CR2 \tag{14}$$

where

$$0 \le \lambda \le 1 \tag{15}$$

Notice that this is a linear interpolative model which incorporates the following highly desirable properties:

- For $\lambda=1$, the proposed model of Eqs. (13), (14) reduces to the standard result of Eq. (9),(10).
- For $\lambda=0$, the proposed model of Eqs. (13), (14) reduces to the standard result of Eq. (11),(12).
- For any λ such that Eq. (15) is satisfied, the proposed model of Eqs. (13), (14) results in quantities that have all the requisite properties to be probabilities (i.e.,

$$0 \leq P'_{fa,eff} \leq 1 \quad (16)$$

$$0 \leq P'_{d,eff} \leq 1 \quad (17)$$

as a natural consequence of the same inequality being satisfied by $P_{d,manual}$; $P_{d,CR2}$; $P_{fa,manual}$; and $P_{fa,CR2}$).

- The combined use of both failure detection techniques corresponding to this proposed linear model may be implemented easily by using the linear weightings in the combined decision rule as discussed below.

A combined decision rule corresponding to the above proposed model is easily obtained. The combined decision rule is as follows:

(1) if

$$(1-\lambda)\cdot \ell_2 + \lambda \cdot \ell_1 \leq (1-\lambda)\cdot K_2 + \lambda \cdot K_1 \quad (18)$$

then choose H_0: no-failure

(2) if

$$(1-\lambda)\cdot \ell_2 + \lambda \cdot \ell_1 \geq (1-\lambda)\cdot K_2 + \lambda \cdot K_1 \quad (19)$$

then choose H_1: a failure (having the critical SNR magnitude) where ℓ_1 is the test statistic and K_1 is the pre-specified decision threshold of the CR2 failure detection technique while ℓ_2 is the test statistic and K_2 is the pre-specified decision threshold of the so-called manual failure detection technique. Notice that for the two extremes of $\lambda=0$ or $\lambda=1$, the decision rule of Eqs. (18) and (19) reduces to the usual decision rules for each separate detection technique.

This proposed model offers an additional appeal when the so-called Receiver Operating Characteristics (ROC) of Fig. 9, which underly all binary decision tests (Ref. 7), are considered. The two hypothetical solid curves in Fig. 9 represent the complete operating capability ranges of the two failure detection techniques. For each of the failure detectors, the usual practice is to trade-off P_d versus P_{fa} by deciding upon a compromise operating point (usually chosen to be in the vicinity of the "knee" of the curve). This operating point is fixed by fixing the decision threshold. The operating points of the two curves are represented by OP and OP' and are denoted by the corresponding two coordinates (P_{fa}, P_d). In general, the plane cannot be ranked (only the real line R may be ordered); so, in general, it is impossible to say, unequivocally, that one operating point is better than another. In Fig. 9, OP has a desirably lower P_{fa} while OP' has a desirably higher P_d. Use of the proposed model of Eqs. (13), (14), (15) would result in an effective OP" somewhere on the dotted line between OP and OP'. For λ=0.5, the effective OP" would be the mid-point of this line segment. The selection of the preferred weighting λ and consequently the associated operation point is discussed in Section 3.5.

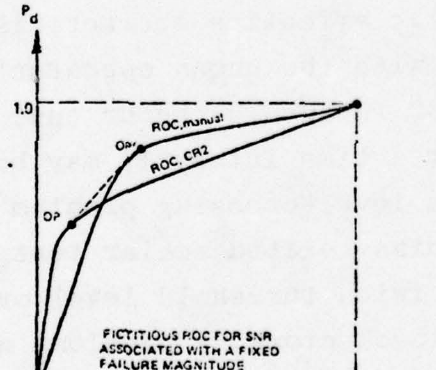


Figure 9 Representative ROC Underlying all Binary Detection Tests

3.4 USE OF OPERATOR TO INTERPRET RESULTS OF FAILURE DETECTION TESTS

Even though the decision rules of the INS2 failure detector, as presented in Section 3.3, are instantaneous rules, it is a common practice to allow a human operator to intervene in the interpretation of the outcome of the test before a switch is thrown. This procedure is followed for conservatism to exploit the useful adaptive capability of a human operator to pick out strong trends over a time interval and, in effect, to create a test with memory, rather than to let an automatic instantaneous computerized test, which could be subject to spurious noise transients, control the throwing of a switch. Use of the human operator in this way would change the effective P_d and P_{fa} characteristics of switch S' from what would be dictated by considering only the instantaneous P_d and P_{fa} characteristics of the CR2 INS2 failure detector. However, once the effective P_d and P_{fa} of the human operator interpreting the results of the CR2 test over a time interval are evaluated, they may be substituted for P_d, CR2 and P_{fa}, CR2 in the availability analysis of Sections 3.1 and 3.2.

More realistic effective characteristic P_d and P_{fa}, which are associated with the human operator's use of the instantaneous CR2 INS2 failure detector test results to recognize trends over a time interval, may be obtained by solving an associated level-crossing problem (i.e., the crossing of the Gaussian-related scalar test statistic above the deterministic decision threshold level over a time interval). Usually level-crossing problems are not very tractable (Ref. 11); however, these interval probability calculations are presently being pursued with an analytically tractable upper bound (Ref. 12) which can be optimized to be as tight as possible to the objective interval probabilities

through the vehicle of a standard quadratic programming problem (Ref. 13, 14). The quadratic programming problem arises naturally in a discrete-time mechanization of the upper bound.

The close connection between level-crossing problems and the test statistic crossing above a decision threshold was also recognized in Ref. 15, where a completely different failure detection approach was investigated. However, only the calculation of the expected level-crossing times were attempted there.

3.5 MECHANICS OF OPTIMAL POLICY SELECTION BASED ON ASYMPTOTIC AVAILABILITY

It is well-known that the reliability function (without repair) asymptotically goes to zero with increasing time (Ref. 2). It is also well-known that the availability function asymptotically goes to some constant positive value as time increases (Ref. 2). Applying this principle to $A_{1/2}(t)^*$ of Eq. (8) with the switch S' characterized by the proposed model of Eqs. (13), (14), (15) allows the evaluation of the asymptotic levels of $A_{1/2}(t)$ over a range of λ weightings between 0 and 1. The value of λ that yields the largest asymptotic $A_{1/2}(t)$ would be chosen and used in the decision rule of Eqs. (18), (19) as the optimal interfacing of the two techniques.

---

*$A_{1/2}(t)$ rather than $A_{1/1}(t)$ was chosen since it represents the principle objective of operating in the primary mode. The methodology is equally applicable to $A_{1/1}'(t)$ however and it is conceivable that $A_{1/1}'(t)$ would be used for the evaluation when interest is in having <u>any</u> good navigation information available.

4. SUMMARY AND CONCLUDING DISCUSSION

The principal objective was to specify and evaluate an INS2 failure detection method. The CR2 failure detection method was specified and its performance was first evaluated theoretically, then demonstrated through simulations. Finally, the CR2 failure detection method was used on real system data to demonstrate that the algorithm is robust enough to handle real world/model mismatches without false alarming while still maintaining the ability to detect failures that occur. The performance of the CR2 failure detection method on real data was consistent with the theoretical predictions. An abridged view of the CR2 failure detection technique and the milestone of achieving satisfactory performance with real data were given in Section 2.

Upon completion of the primary objective, it was then necessary to specify how the CR2 INS2 failure detector* should interface with the entire system, including any additional INS2 failure detection aids (manual or automatic) that are subsequently developed to help the navigation operator decide in what mode he should operate the system. It is also necessary to evaluate how the CR2 failure detector affects the whole three-INS complex. A rigorous theoretical framework was detailed in Section 3 for using the standard techniques of reliability theory to determine the effect of the characteristic P_d and P_{fa} of the CR2 failure detector on the

*Method applicable to any INS2 failure detection method that uses the same filter inputs as used by the CR2 test as long as the characterizing P_d and P_{fa} have been evaluated.

overall asymptotic availability of the three-INS complex. When interfacing with another INS2 failure detection method is required, the proposed new model of the effective P_d and P_{fa} characteristics associated with the switch, used as a standard model for human operator decisions, may be used to determine the optimum relative weightings between the two methods that achieves the highest asymptotic availability. The same relative weightings are also inherited by the two INS2 failure detection techniques in a joint implementation as discussed in Section 3.3.

When only one INS2 failure detector is used, the proposed new model for effective P_d and P_{fa} (for switch S' in mode selection) reduce to the standard model without controversy. Presently, there appears to be a scarcity of ROC data on human operator decisions on master INS1 selection. Hopefully, this void will be filled in the future through the compilation of adequate performance data and/or statistical design of experiments to quantitatively evaluate the effect of man-in-the-loop. The theoretical framework for quantitatively evaluating the effect of the CR2 INS2 failure detector has been completely worked out, but is waiting on the specification of P_d, P_{fa} characteristics for switch S since this switch (operator decision) is depended upon in both the primary and back-up mode of the three-INS complex and availability results obtained by proceeding without this critical information may be very misleading.

One methodology, recently proposed for modeling man-machine availability as an allocation problem (Ref. 16), allows several levels of complexity and additional realism such as:

- grade level and years of experience of each repairman
- number of repairmen assigned to each repair action
- time spent by each repairman on the repair .

However, dynamic programming is required to solve even the most simple non-degenerate problems within this framework. In spite of some relatively recent theoretical strides in the area of implementing dynamic programming algorithms (Ref. 17, 18, 19) to reduce the so-called "curse of dimensionality" associated with dynamic programming problems (Ref. 20), the implementation of a dynamic programming algorithm still remains a rather formidable problem and usually requires a large computer allotment. In contrast, the main computation required for the availability analysis of Section 3 is a relatively minor iteration of a sparse, constant coefficient, linear difference equation of medium order. A more detailed model may be obtained, using the technique of Section 3, by modeling more elements.

Theoretical strides have been made recently in applying the techniques of modern control to the problem of modeling the effect of the operator-in-the-loop on the performance of the overall system. Two different approaches which both make use of Kalman filters are:

(1) human operator modeling within the theoretical framework of an optimal stochastic regulator (Ref. 21, 22, 23)

(2) modeling of human operator remnants using maximum likelihood methods for parameter identification (Ref. 24, 25).

It is anticipated that these new modeling approaches will yield more accurate final evaluations of the effective P_d and P_{fa} of the human operators.

The methodology discussed in this paper may also have potential application to those USAF situations in which a human operator must decide between two or more INS or navigation aids such as in:

- the C-141 dual IMU installation
- the C-5 planned retrofit
- the B-1 .

REFERENCES

1. Nash, R.A., Kasper, J.F., et al, "Applications of Optimal Smoothing to the Testing and Evaluation of Inertial Navigation Systems and Components," IEEE Transactions on Automatic Control, Vol. AC-16, No. 6, 1971.

2. Shooman, M.L., Probabilistic Reliability: An Engineering Approach, McGraw-Hill Book Company, New York, 1968.

3. Kerr, T.H., "A Two Ellipsoid Overlap Test for Real-Time Failure Detection and Isolation by Confidence Regions," Proceedings of the Fifth Annual Pittsburgh Conference on Modeling and Simulation, Part 2, p.1003, 1974, (a short working paper).

4. Kerr, T.H., "A Two Ellispoid Overlap Test for Real-Time Failure Detection and Isolation by Confidence Regions," Proceedings of the Conference on Decision and Control, pp.735-742, 1974.

5. Kerr. T.H., "Real-Time Failure Detection: A Static Nonlinear Optimization Problem That Yields a Two Ellipsoid Overlap Test," to appear in the Journal of Optimization Theory and Applications.

6. Kerr, T.H., "Applying Stochastic Integral Equations to Solve a Particular Stochastic Modeling Problem," Ph.D. Thesis, University of Iowa, 1971.

7. Van Trees, H.L., Detection, Estimation and Modulation Theory; Part I: Detection, Estimation, and Linear Modulation Theory, John Wiley and Sons, Inc., New York, 1968.

8. Schweppe, F.C., Uncertain Dynamic Systems, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1973.

9. Hartman, P.D., "On the Theory of Confidence Set Estimation and Detection," Ph.D. Thesis, Polytechnic Institute of Brooklyn, 1972.

10. Nakagawa, T. and Osaki, S., "Stochastic Behavior of 2-Unit Standby Redundant Systems with Imperfect Switch-Over," IEEE Transactions on Reliability, Vol. R-24, No. 2, June 1975.

11. Wong, E., "Recent Progress in Stochastic Processes - A Survey," IEEE Transactions on Information Theory, Vol. IT-19, No. 3, May 1973.

12. Gallager, R.G., and Helstrom, C.W., "A Bound on the Probability that a Gaussian Process Exceeds a Given Function," IEEE Transactions on Information Theory, Vol. IT-15, No. 1, January 1969.

13. Boot, J.C.G., Quadratic Programming: Algorithms-Anomalies-Applications, Rand McNally and Company, Chicago, Illinois, 1964.

14. Fletcher, R., "A FORTRAN Subroutine for General Quadratic Programming," Harwell Report, No. AERE.6370, United Kingdom Atomic Energy Research Group, June 1970.

15. Hines. W.G.S., "A Simple Monitor of a System with Sudden Parameter Changes," IEEE Transactions on Information Theory, Vol. IT-22, No. 2, March 1976.

16. LaSala, K.P., Siegal, A.I., et al, "Allocation of Man-Machine Reliability," Proceedings of 1976 Annual Reliability and Maintainability Symposium, Las Vegas, Nevada, 1976.

17. Larson, R.E., State Increment Dynamic Programming, American Elsevier Press, New York, 1967 .

18. Larson, R.E. and Korsak, A.J., "A Dynamic Programming Successive Approximations Technique with Convergence Proofs," Automatica, Vol. 6, No. 2, March 1970.

19. Bertsekas, D.P., Dynamic Programming and Stochastic Control, Academic Press, New York, 1976.

REFERENCES (Cont'd)

20. Bellman, R., Introduction to Matrix Analysis, McGraw-Hill Book Company, New York, 1970.

21. Kleinman, D.L., Baron, S., et al, "An Optimal Control Model of Human Response, Part I: Theory and Validation," Automatica, Vol. 6, 1970.

22. Kleinman, D.L., Perkins, T.R., "Modeling Human Performance in a Time-Varying Aircraft Tracking Loop," IEEE Transactions on Automatic Control, Vol. AC-19, No. 4, August 1974.

23. Broussard, J.R., Stengel, R.F., "Stability of the Pilot Aircraft System in Maneuvering Flight," to appear in Proceedings of the Twelfth Annual Conference on Manual Control, University of Illinois at Urbana-Champaign, May 25-27, 1976.

24. Sheridan, T.B., Ferrell, W.R., Man Machine Systems, MIT Press, Cambridge, Mass., 1974.

25. Repperger, D.W., Junker, A.M., A Maximum Likelihood Approach for Identifying Human Operator Remnant in a Tracking Task," Proceedings of the Conference on Decision and Control, Houston, Texas, 1975.

ANALYTICAL MODELS OF MAITNENANCE DECISION PROCESSES

by

W. Stephen Demmy
Wright State Univeristy

This paper presents a series of analytical models of maintenance decision processes. In the context of a given maintenance system, these models should be useful in answering questions such as:

1. What are the key maintenance decisions?
2. How can we improve these decisions?
3. What is the expected impact of a specific change in one area upon the entire system?

In this section, we briefly sketch the characteristics of the logistics support systems for the KT-73 Inertial Measurement Unit, a major component of the navigation system for A-7D and A-7E aircraft. As we shall see, this system is extremely complex. This complexity is not unique to the KT-73 logistics system; many other Air Force items have logistics systems of similar or greater complexity.

Our ultimate objective is to develop tools for optimizing the decision processes of such complex networks. In working toward this goal, Section II reviews the current state of the art for optimizing maintenance decision processes. Next, we discuss the features of a series of increasingly complex maintenance decision systems. Section III discusses the major factors involved in maintenance decisions for single tests, while Section IV presents methods for setting optimum cutting scores for such tests. Section V generalizes the results of Section IV to consider tests in which multiple readings are obtained; for example, a single test of an inertial measurement unit may provide readings three separate slew rates; one reading for each of the X, Y, and Z coordinates. Methods for determining optimum multivariate cutting scores are presented in this section.

In Section VI, networks of test and repair facilities are considered, and a method for computing optimal cutting scores for the system of tests is presented.

In most systems, the detailed data required to apply the tools developed in Sections III through VI will not be readily available, but

may be obtained using experimental design and cost analysis techniques described in the main body of this report. Applying these techniques can be expensive; what is needed is a tool for evaluating the potential savings that may be obtained using these measurement techniques. Hence, Section VII discusses a procedure for identifying critical test or repair activities, i.e., activities that offer high potential paybacks if current test and/or repair errors are reduced.

Let us now consider the characteristics of the KT-73 logistics support system.

THE KT-73 INERTIAL MEASUREMENT UNIT

The KT-73 Inertial Measurement Unit (IMU) is a major component of the navigation system installed in A7D and A7E aircraft. At present, the Aerospace Guidance and Metrology Center (AGMC) is the only depot repair facility for this IMU.

The KT-73 IMU is basically a gyroscopically stabilized "platform" upon which are mounted electro-mechanical accelerometers that sense and measure inertia variances associated with movement of the unit. During flight, the external case of the IMU pitches and rolls with the aircraft to which it is attached; but inside the IMU, the gyroscopically stabilized platform remains in the same constant spatial-attitude regardless of the movements of the external case. Since this stabilized platform presents an essentially "fixed" attitude reference, the accelerometers mounted on it can provide accurate data concerning acceleration of the IMU (and therefore the aircraft) along any vector (direction) from a chosen reference point. By comparing these measurements with elapsed time, the IMU can determine distance and direction information that provides a basis for navigation of the aircraft.

The basic physical construction of the KT-73 IMU is shown in a simplified form in Figure 1. From the figure it can be seen that the IMU may be considered to be basically composed of three "nested" assemblies. The gyroscopically stabilized platform (<u>not</u> visible in Figure 1) is

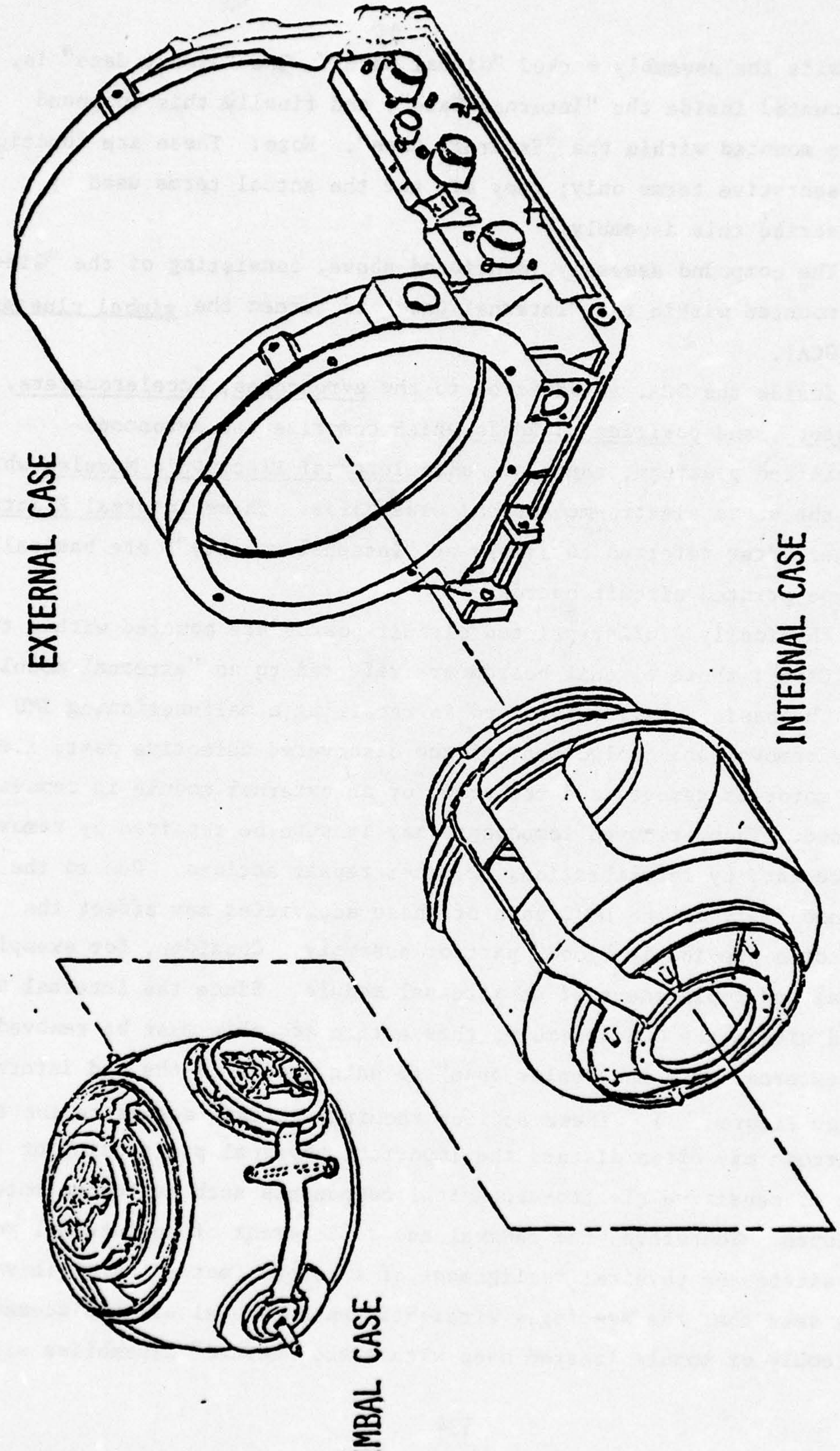


FIGURE 7. KT-73 IMU ASSEMBLY

located inside the assembly marked "Gimbal Case". The "Gimbal Case" is, in turn, mounted inside the "Internal Case"; and finally this compound assembly is mounted within the "External Case". Note: These are functionally representative terms only; they are not the actual terms used by AGMC to describe this assembly.

The compound assembly, mentioned above, consisting of the "Gimbal Case" mounted within the "Internal Case" is termed the gimbal cluster assembly (GCA).

Inside the GCA, in addition to the gyroscopes, accelerometers, torque motors, and position pickoffs which comprise the gyroscopically stabilized platform, there are also Internal Electronic Modules which interface the above electro-mechanical assemblies. These Internal Electronic Modules (hereafter referred to simply as "internal modules") are basically plug-in type printed circuit boards.

Physically similar printed circuit boards are mounted within the "External Case"; these circuit boards are referred to as "external modules".

The basic activity involved in repairing a malfunctioning IMU is simply the removal and replacement of the discovered defective part, i.e., a torque motor is removed and replaced, or an external module is removed and replaced. These removed components may in turn be repaired by removal and replacement, by recalibration, or other repair actions. Due to the extreme complexity of the IMU, each of these activities may affect the operation of a previously "good" part or assembly. Consider, for example, the removal and replacement of an internal module. Since the internal module is located within the GCA assembly, this entire assembly must be removed from the external case and "split open" to gain access to the bad internal module (see Figure 1). These actions required to gain access to the electronic circuit may often disturb the important physical positioning or alignment of sensitive electro-mechanical components such as torque motors or gyroscopes. Therefore, the removal and replacement of an internal module may necessitate the physical realignment of a torque motor. From this example it can be seen that the seemingly straightforward removal and replacement of an assembly or module located deep within the "nested" assemblies will

necessitate multiple disassembling to gain accessability, which in turn may require the physical realignment of many electro-mechanical components.

On the other hand, removal and replacement of a malfunctioning external module would not require the other internal assemblies to be disturbed in an effort to gain access to the defective part. Therefore the actual removal and replacement of the defective part would be all that would be required. Throughout the remainder of this research, a distinction will be drawn between <u>external</u> and <u>internal</u> repair. External repair will refer to repair activities associated with External Case components and internal repair will designate repair activities associated with parts located within the "nested" internal assemblies.

THE KT-73 LOGISTICS SYSTEM

The basic components of the maintenance and repair system for KT-73 IMU's is illustrated in Figure 2. Block 1 represents a KT-73 that is performing its primary function; namely, assistance in the navigation of A7-D or A7-E aircraft.

As illustrated by Block 2, built-in-test equipment is used to monitor key performance measures of the IMU while the unit is still installed in the aircraft. If these tests indicate the unit is malfunctioning, the IMU is removed from the aircraft, and moved to the field maintenance activity (Block 3) for further testing.

Field maintenance activities consist of more precise testing and diagnosis than can be performed while the equipment is on the aircraft. If these tests indicate a unit is bad, it is repaired and retested in the field, if possible (Block 4). When the field tests indicate the item is good, it is installed in the aircraft and is ready for pre-flight checkout. On the other hand, if field repair is not possible or appears undesirable, the unit is shipped to the depot (Block 5).

At present, the Aerospace Guidance and Metrology Center (AGMC) is the only depot repair location for KT-73 Inertial Measurement Units. As illustrated in Figure 2, the major aspects of depot level repair at AGMC include:

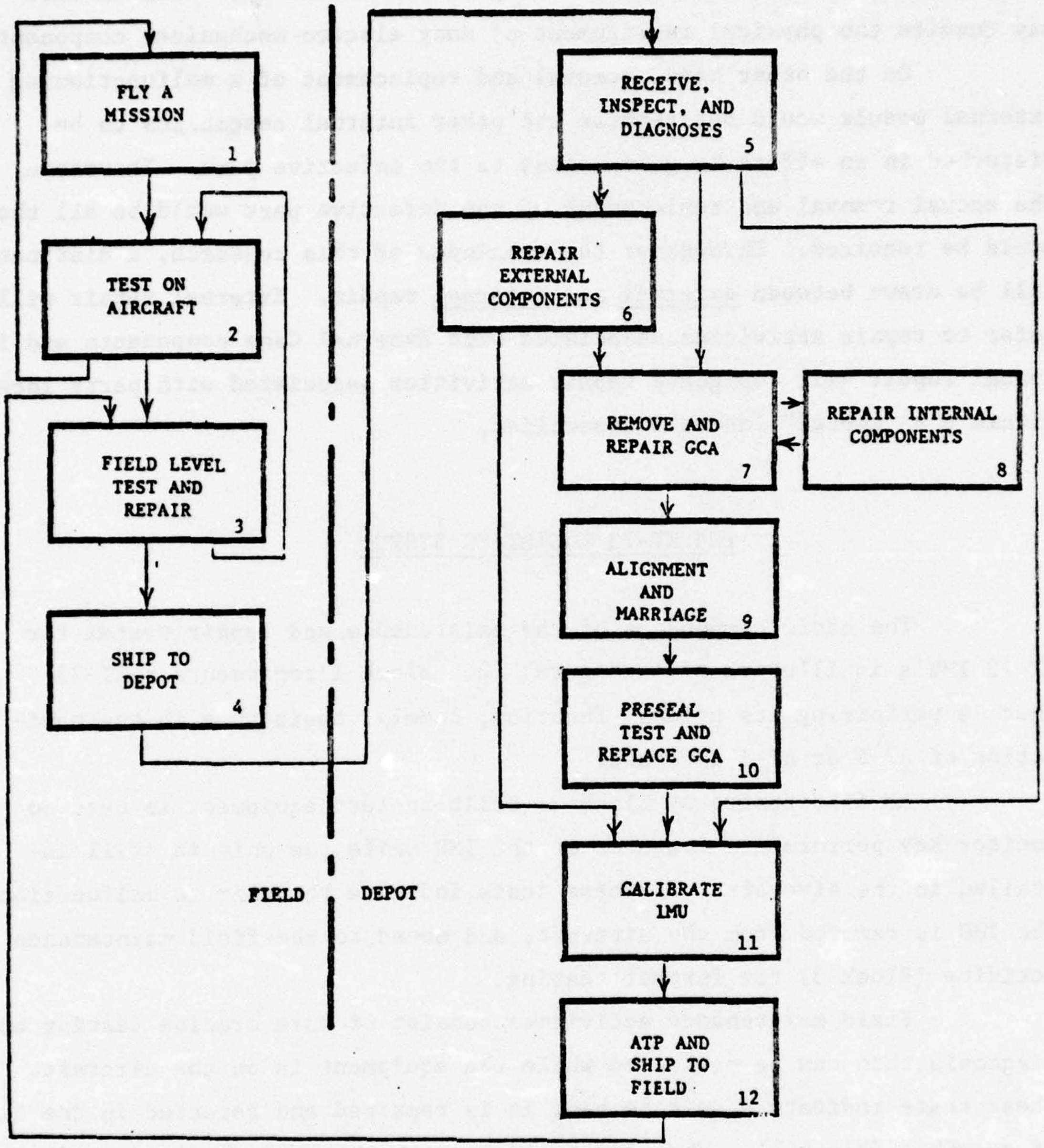


FIGURE 2. KT-73 IMU MAINTENANCE PROCESS FLOW CHART

a. Receiving, inspection, and diagnosis of incoming units.

b. Repair or replacement of components that are considered faulty.

c. Calibration and final testing of the overhauled unit.

d. Shipment to the field installations.

These activities are illustrated in Blocks 6 through 12 of Figure 2.

A more detailed flow chart and associated cost data for KT-73 depot repair activities is presented in Figure 3. This data was developed by Watson and Waterman (1974) as part of a cost analysis of the KT-73 repair process. A majority of the following discussion is based on their work.

In examining Figure 3 it should be recognized that each of the 16 major "stages" shown actually represents one or more activities. The stage shown as Repair and Replace External Module, for example, actually covers repair or replacement of four different modules, as well as a test to ensure that the fault identified has actually been repaired. Consequently, although 16 stages are shown, the work performed at each of these stages actually covers correction of a range of specific malfunctions, and does not represent a set of "standard" tasks.

After a KT-73 IMU is received at AGMC, and the necessary paperwork is completed, the IMU is forwarded to the KT-73 Repair Shop. All IMU's arriving at the shop from the field are then processed through the Receiving Stage. During this stage a complete check of the IMU is performed using an automated test station, to confirm the malfunction reported by the previous user and to identify any additional problem areas.

Dependent upon the results of the receipt tests, the IMU is classified as requiring either internal (cluster) repair or repair of one of the external modules. If internal repair is required, the IMU is sent to one of the stages at the next level, shown on Figure 3 as Repair and Replace Internal Module, Repair and Replace Torque Motor or Electronics Repair stages. If external repair is required, the IMU is sent to Repair and Replace External Module Stage. The necessary repair is then performed and the IMU is tested to determine whether the repair was successful.

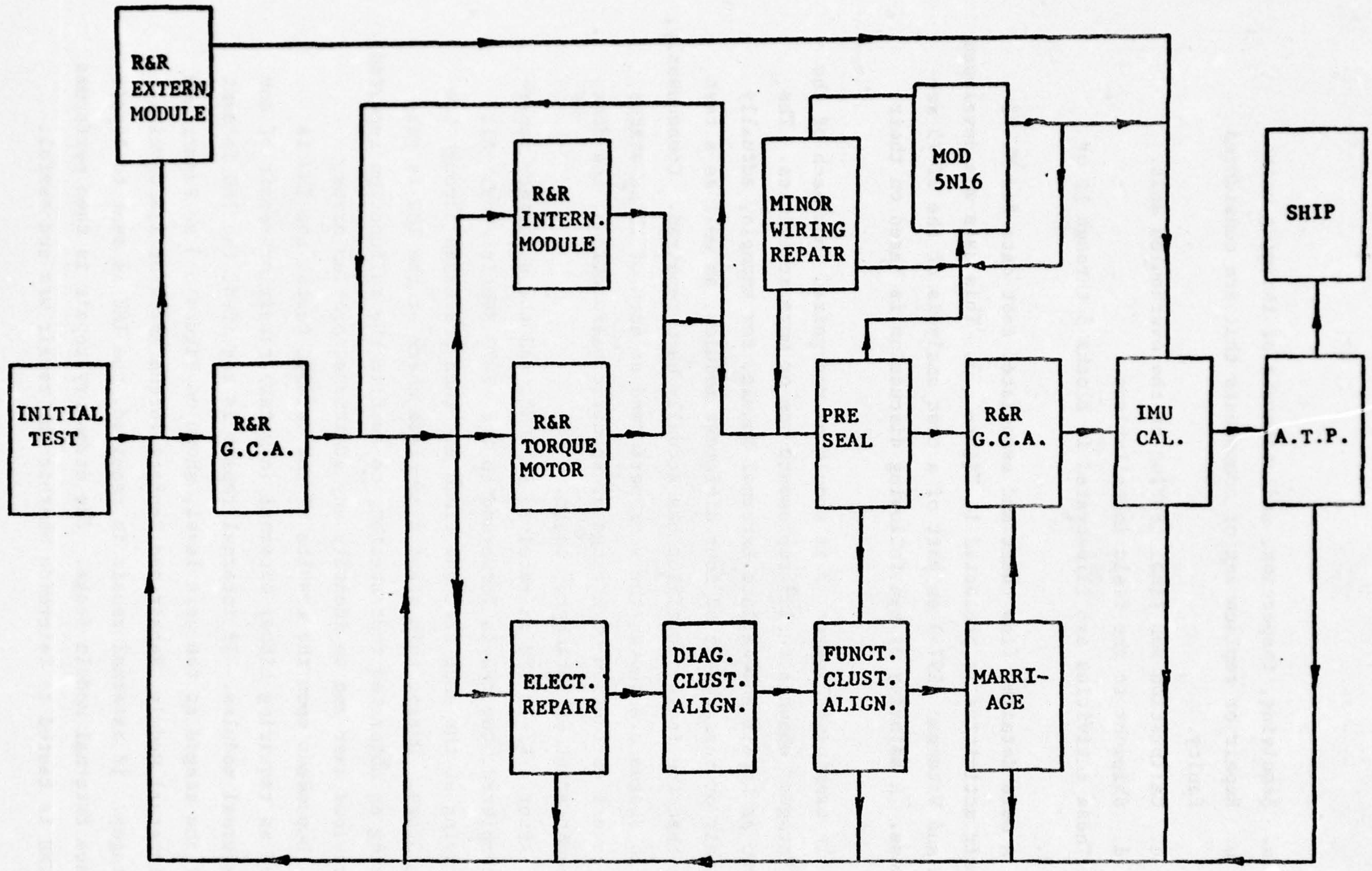


FIGURE 3. SIMPLIFIED FLOWCHART - KT-73 REPAIR PROCESS

Provided that the results of the test show that the original malfunction has been corrected, the IMU then moves to the next stage in the repair sequence. The actual repair sequence for an IMU is dependent upon the faults present in a particular unit, and consequently all IMU's do not follow the same path through the process. From Figure 3 it may be seen that the number of paths which a unit may follow is quite large, depending on the actual faults present. This progression from stage to stage in the process continues until the IMU reaches the final test stages.

Final test and calibration of a repaired IMU is carried out in the IMU Calibration and Final Test stages, and as a result of the tests conducted during these stages the IMU is classified either as "satisfactory" for return to service, or as "unsatisfactory" and requiring further repair. Each of these stages comprises a number of specific functional tests. The Final Test stage, for example, includes 13 individual tests, each designed to evaluate the performance of the IMU in a particular phase of its operation. Successful completion of all 13 tests is required before the IMU is classified as satisfactory and shipped to field activities.

Ideally, items would progress sequentially from one stage to the next until the IMU reaches the final test stage and is shipped to the field. Unfortunately, tests at a given stage may uncover problems that were introduced earlier in the repair process; such faulty items must be returned to the earlier stage for correction. This "feedback" can be very expensive.

Consider, for example, the case where an IMU which has been repaired at the Internal Module Stage reaches the Final Test Stage and is found to be unsatisfactory. If the unsatisfactory condition can be directly diagnosed as requiring work at the Internal Module Stage, the IMU must then be "fed back" to that stage for correction of the fault. However, because of the error early in the process, the unit must repeat not only the Internal Module and Final Test Stages, but all intervening stages in the process.

While this type of feedback is costly enough, next consider the same situation, but assume the fault is improperly diagnosed, and the unit is incorrectly sent to the Remove and Replace Torque Motor Stage. After

work is performed at this stage, the unit must still eventually be returned to the Internal Module Stage to correct the true cause of the fault.

If more than one repair or diagnosis error occurs, a number of stages may have to be repeated more than once, with a corresponding increase in cost.

The above discussion illustrates the great complexity of the KT-73 test and repair system. Let us now review the current state of the art for optimizing decision processes in such complex systems.

II. REVIEW OF THE LITERATURE

A critical factor in the design of any maintenance system is the precision of the diagnosis process. If an automobile has a flat tire, a decision to change that tire is almost certainly correct. However, if an inertial measurement unit produces large navigation errors, the actions needed to correct the condition are far less obvious. For complex equipment, it is usually very difficult to determine the true cause of a given operational problem. For example, large navigational errors in a KT-73 Inertial Measurement Unit (IMU) may be caused by a faulty A-200 accelerameter; there are several of these in a KT-73 unit. Furthermore, to determine which, if any, of these is faulty requires complete disassembly of the IMU, a very expensive and time consuming process. To make matters worse, it is possible that other components that are originally good may be damaged or misaligned during the disassembly operation.

Many techniques are available for determining cost-effective repair decision when there is perfect diagnosis, but much less is known about correct maintenance policies in the face of diagnostic uncertainty. Several authors have modeled the latter problem as a Markovian decision process with probabilistic observation of states. This includes the work of Astrom (1965), Eckles (1968), Satia and Lave (1973), Smallwood and Sondik (1973), and White (1976). These authors explore the mathematical characteristics of optimal decisions for these processes, and propose computational algorithms for determining optimum decisions; however the algorithms appear useful for only very simple problems. The above authors assume that all transition and observation probabilities are known. In practice, estimation of these probabilities can be a serious problem. A non-linear programming approach for estimating the transition probabilities for unobserved states has been reported by Genet (October, 1970). Unfortunately, the large amount of computation required by this approach has limited its usefulness.

Another approach to decision making with imperfect information is straightforward application of decision and utility theory techniques.

This approach uses decision trees to describe the structure of the decision problem. This structuring involves identification of available alternatives, their possible consequences, and the conditions under which those consequences might occur. Utility theory is then used to develop scales for measuring the relative desirability of actions that lead to several consequences (e.g. medical "repair" actions may lead to consequences such as "cure", "no effect", or "severe complications"). Finally, Bayesian inference methods are used to combine subjective probability estimates with empirical data to determine a "best" action. Application of these techniques to medical diagnosis decisions is reported by Betaque and Gorry (1971). Brown and McAllister (1962) and Firstman and Gluss (1960) discuss use of these techniques for fault diagnosis in electronic equipment.

Each of the technical approaches described above should prove useful in the design of improved diagnosis and repair systems. Markov Process models may be used to identify some (but not all) of the key decision points in a multi-stage system and to estimate the impact of proposed changes. Such a model should be useful in determining which alternatives should be studied in more detail.

The decision theory approach appears to have excellent potential for improving specific diagnostic and repair decision. The theory is well developed; practical application appears to depend on development of data processing tools to implement the calculations and on development of decision trees describing the technology of specific repair processes.

In theory, decision trees could be developed to aid in the design of multi-stage test/repair systems. In practice, however, the extreme complexity of most Air Force equipment will probably limit usefulness of this technique to the design of single diagnosis and repair stages.

In summary, no single technique (or set of techniques) currently exists for determining optimum diagnosis and repair systems. However, tools for attacking pieces of the total problem do exist. In the following sections, we will illustrate the use of these tools in the analysis of specific diagnosis/repair decisions.

III. UNIVARIATE TESTS

Suppose that the condition of a given item can be classified into one of two possible states: operational or failed. If the item is operational, we assume it is capable of performing assigned missions within the item's design specifications; if failed, the item cannot perform to these requirements. For simplicity, we will say the item is "good" if it is operational, and "bad" if it is not.

Now suppose that we cannot directly observe whether the item is good or bad. Rather, we can only estimate the item's condition by performing a series of electrical and mechanical measurements. Suppose the net result of these actions is summarized in a single number, which we will call the "reading". As noted by Genet ("Avionics Cost Reduction Through Improved Tests"), variations in individual readings from test to test are often due to:

a. Variability within the item itself

b. Test station calibration and noise

c. Operator variability.

Variability may also be caused by interactions among these sources or by other causes.

Figure 4 illustrates a situation in which there is no variability in test readings. In this example, a good item always has a reading of 100, while a bad item always has a reading of 50. This is an example of a perfect test; there is a one-to-one correspondence between the test reading and the true state of the item.

Figure 5 illustrates a case in which there is item variability but no other source of variance in the readings. In this case, if we were to take readings on a large number of items that were known to be good, the average reading would still be 100, but specific readings would vary about the mean. Similarly, if a large number of bad items are measured, the average value is 50 in this case, but specific items would have values both above and below the mean due to differences within each item.

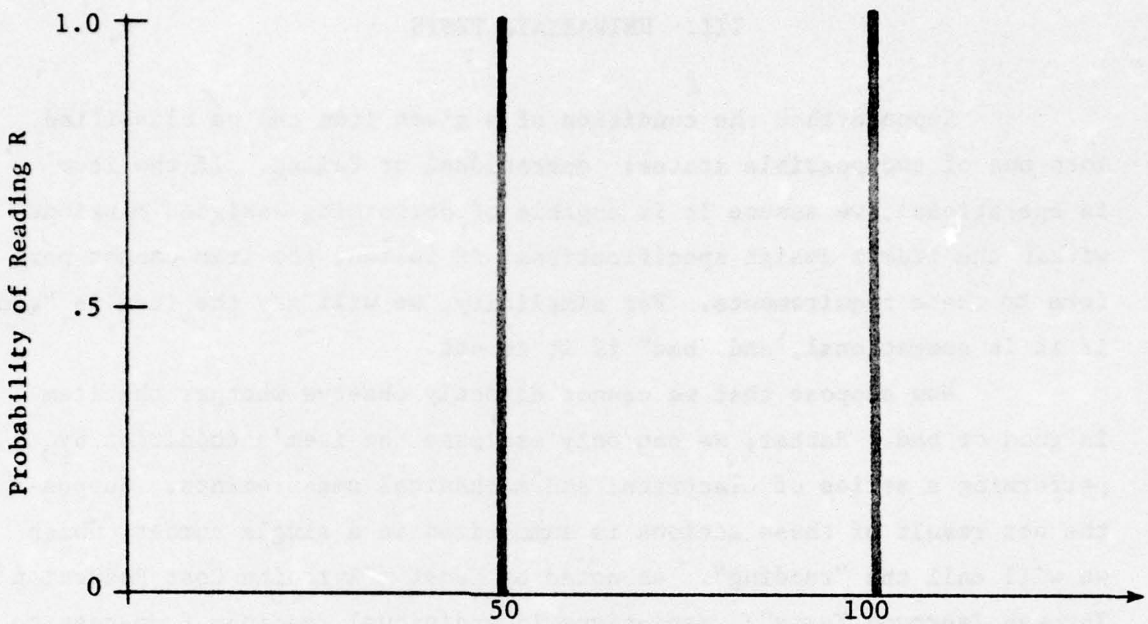


FIGURE 4. READINGS FOR A PERFECT TEST

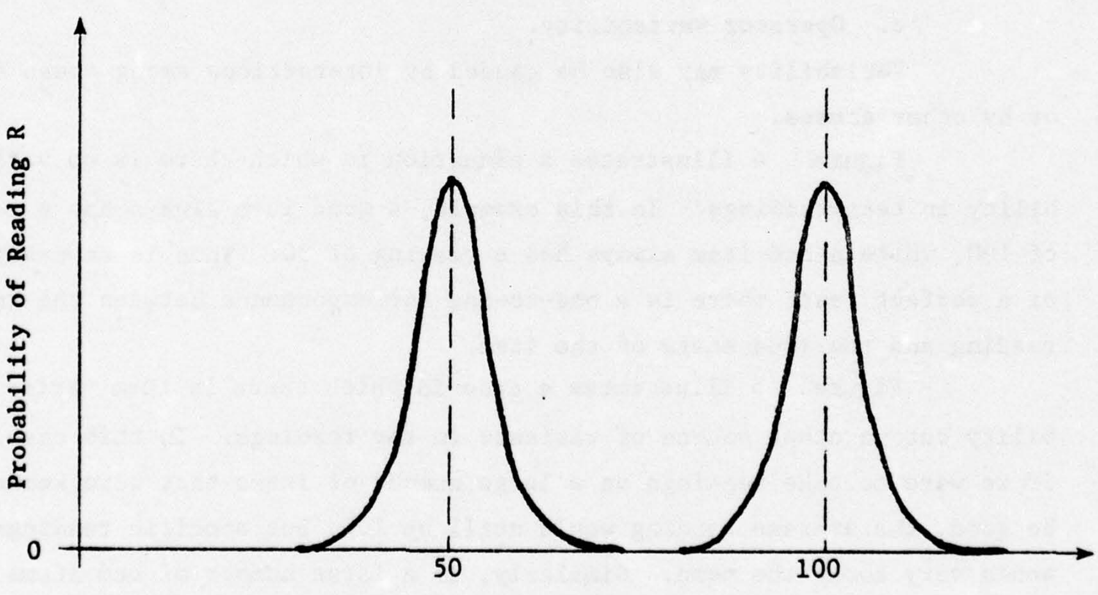


FIGURE 5. READINGS WITH ITEM VARIABILITY AND NO OTHER MEASUREMENT ERRORS

Now let us consider what happens when there are still more sources of reading variability.

Lec i indicate the ith reading in a series of readings, and let

R_i = value read on the test equipment
V_G = average value of a good item
V_B = average value of a bad item
I_i = item error, or "noise", from the "true value" of the item
S_i = station error
O_i = observer error
U_i = unexplained error; i.e. error due to other than item, station, or observer error, including interaction effects among items, stations, and observers.

Then, if an item is good,

$$R_i = V_G + I_i + S_i + O_i + V_i \quad (1)$$

The expected value of the reading, E(R), is then

$$E(R_i) = V_G + E(S_i) + E(O_i) + E(U_i) \quad (2)$$

since $E(I_i)$, the expected value of item error for a good item, is zero by definition of $V_G + I_i$. Similarly, if an item is bad,

$$E(R_i) = V_B + E(S_i) + E(O_i) + E(U_i) \quad (3)$$

If $[E(S_i) + E(O_i) + E(U_i)] = 0$, the test is unbiased, since the expected reading then equals the true mean value. Otherwise, the test is biased. If a test is biased, an unbiased reading may be obtained by simply subtracting the amount of bias out of the reading. This is essentially what happens when equipment is calibrated. In our remaining discussions, we assume that specific tests are unbiased.

Let σ_S^2, σ_O^2, σ_U^2 denote respectively the variance due to station, observer, and unexplained error sources. Also, let σ_G^2 and σ_B^2 denote the respective items variances of good and bad units. Since σ_U^2 includes all unexplained variation, including interactions effects, for good items

$$\sigma_R^2 = \sigma_G^2 + \sigma_S^2 + \sigma_O^2 + \sigma_U^2 \tag{4}$$

while for bad items

$$\sigma_R^2 = \sigma_B^2 + \sigma_S^2 + \sigma_O^2 + \sigma_U^2 \tag{5}$$

Hence, if we know the average values and variances for good and bad items (V_G, σ_G^2, and V_B, σ_B^2, respectively) and the sum of variances $\sigma_S^2 + \sigma_O^2 + \sigma_U^2$ due to station, operator, and *unexplained error sources*, we can compute the mean and variance of reading errors for good and bad items. Note that we do not need to know the individual components σ_S^2, σ_O^2, or σ_U^2; only the sum is needed. To determine the variance of the sum is a much easier estimation problem than estimating the characteristics of each variance component individually.

If each of the error sources are normally distributed, specific readings will also be normally distributed with the parameters given by (2) - (5) above. Figure 6 illustrates how Figure 5 might change if additional error sources are present. With more error sources, the variance is higher, and hence the distribution of readings is less peaked; there is also a wider range for possible observations.

Note that the distribution of readings for good and bad items overlap in Figure 6. For example, some good items can have readings of less than 80, while some bad items can have readings of more than 80. Here we no longer have a perfect test. There is no longer a one-to-one correspondence between a reading and a good or bad item. However, the test is still useful for distinguishing between good and bad items, since the area of overlap is very small.

Figure 7 illustrates a case where the distribution of readings overlap significantly. In this case, a specific reading tells us very little about the true state of the item being tested. If the two distributions overlap completely, the reading tells us nothing about the item's true condition. Such a test is said to be invalid.

CUTTING SCORES AND OPERATING CHARACTERISTIC CURVES

One method of standardizing test decision making is through the use of cutting scores. For example, suppose that for the item illustrated in Figure 6, a cutting score of 80 is established. In testing a given item, a specific reading R will be obtained. If this reading is greater than the cutting score, 80, the item passes the test; and is then treated as if it were a good item. If the reading is less than 80, the item would be assumed to be bad, and appropriate repair actions would then be taken. Note from Figure 6 that with a cutting score of 80, some bad items will pass the test, and some good ones will fail. On the other hand, if the cutting score is set at 100, no bad items will pass the test, but many good items will fail it.

One way to quantify the sensitivity of a given item to different cutting scores is shown in Figure 8. This figure plots the probability of passing the test versus various cutting scores. Two curves are shown: one for items that are in fact good, another for bad items. We shall refer to these and similar curves as the characteristic curves, or C-curves, for the test, since the curves characterize the power of the test to discriminate between good and bad items. In constructing Figure 8, we assumed that $V_G = 100$; $\sigma_G = 10$; $V_B = 50$; $\sigma_B = 15$. This is the same data used to construct Figure -6.

Note that with a cutting score of 80, there is a 2.3% chance of passing a bad item, while there is also a 2.3% chance of failing a good item. On the other hand, with a cutting score of 100 no bad ones will pass the test; but 50% of all good items will fail it.

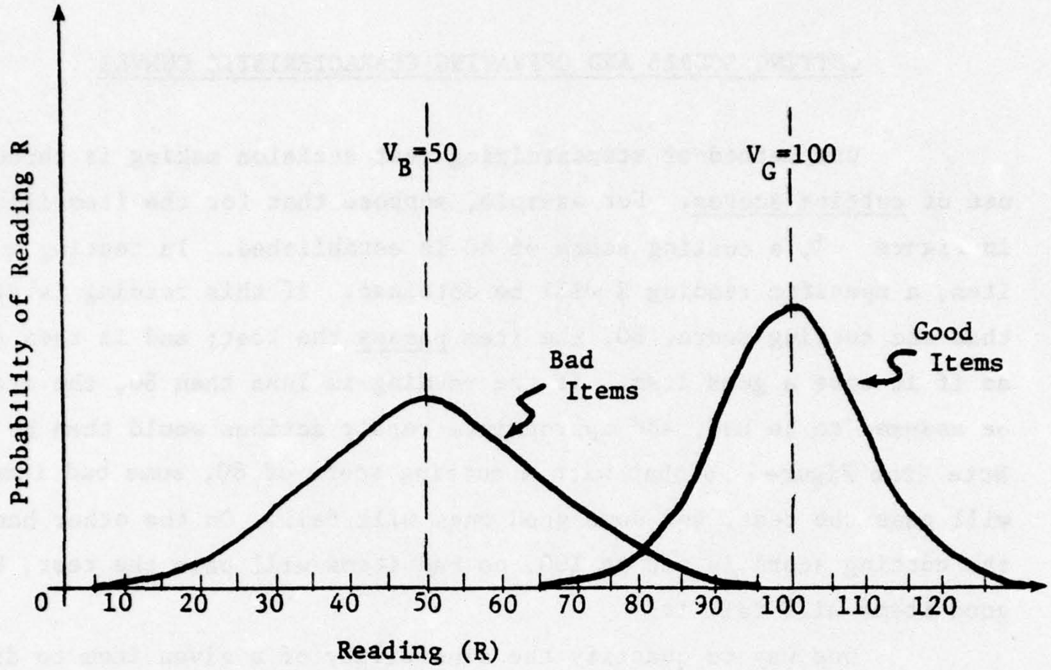


FIGURE 6. DISTRIBUTION OF READINGS WITH SERVERAL ERRORS

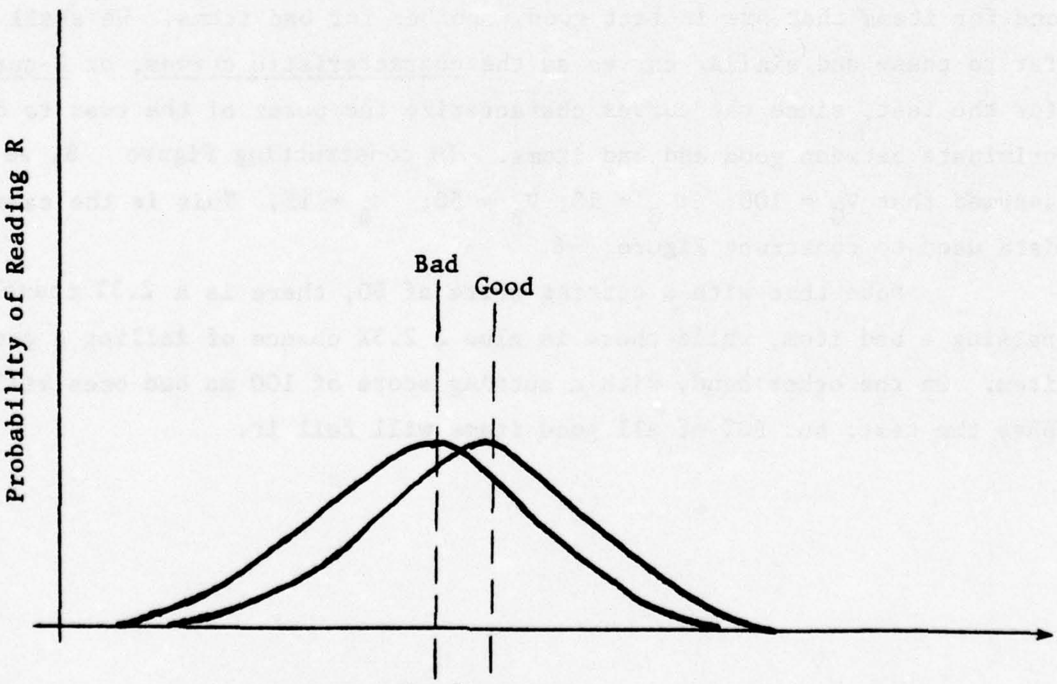


FIGURE 7. ILLUSTRATION OF AN INVALID TEST

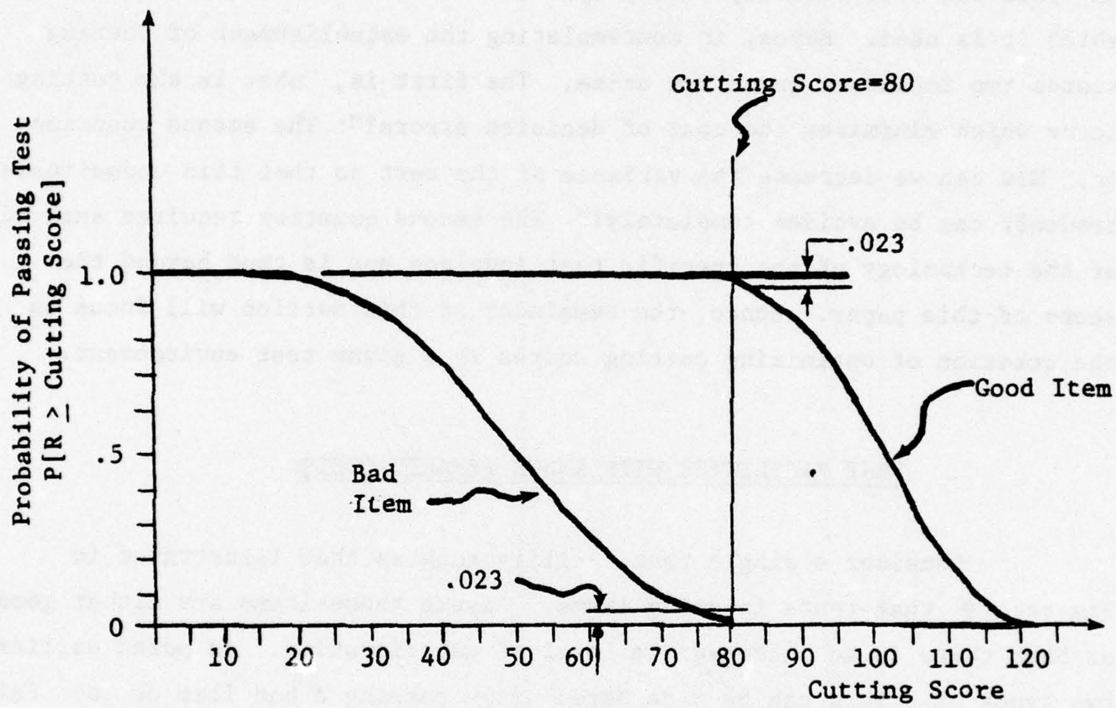


FIGURE 8. THE RELATIONSHIP BETWEEN CUTTING SCORE AND THE PROBABILITY OF PASSING A TEST

IV. OPTIMUM CUTTING SCORES

In setting cutting scores, two types of costs are incurred. If the cutting score is set too high, some good items will fail the test, and will mistakenly be sent to undergo repair or overhaul actions. On the other hand, if the cutting score is set too low, some bad items may pass the test, and may thus jeopardize the success of any mission in which it is used. Hence, in comtemplating the establishment of cutting scores two important questions arise. The first is, "what is the cutting score which minimizes the cost of decision errors?" The second question is, "How can we decrease the variance of the test so that this undesirable tradeoff can be avoided completely?" The second question requires analysis of the technology of the specific test involved and is thus beyond the scope of this paper. Hence, the remainder of this section will focus on the question of optimizing cutting scores in a given test environment.

TEST FACILITIES WITH KNOWN PENALTY COSTS

Consider a single test facility such as that illustrated in Figure 9 that tests incoming items. Assume these items are either good or bad; there is no intermediate level of deterioration. As noted earlier, two types of errors can be made here: (a) passing a bad item or (b) failing or rejecting an item that is, in fact, good. Suppose there are known economic penalties associated with each of these errors; specifically, let

CRG = cost of rejecting a good item

CAB = cost of accepting a bad item.

For example, if a good item fails a field level test, it may be sent to the depot for repair. CRG then represents the cost of transporting the item to and from the depot plus the cost of work performed at the depot level. On the other hand, if a bad item passes the test the success of subsequent missions may be jeopardized. CAB should then represent an economic measure of the cost of a jeopardized mission. [Another approach is to treat CAB as a parameter, and then develop cost/effectiveness curves that relate probability of mission success to logistics support costs.] If passing a

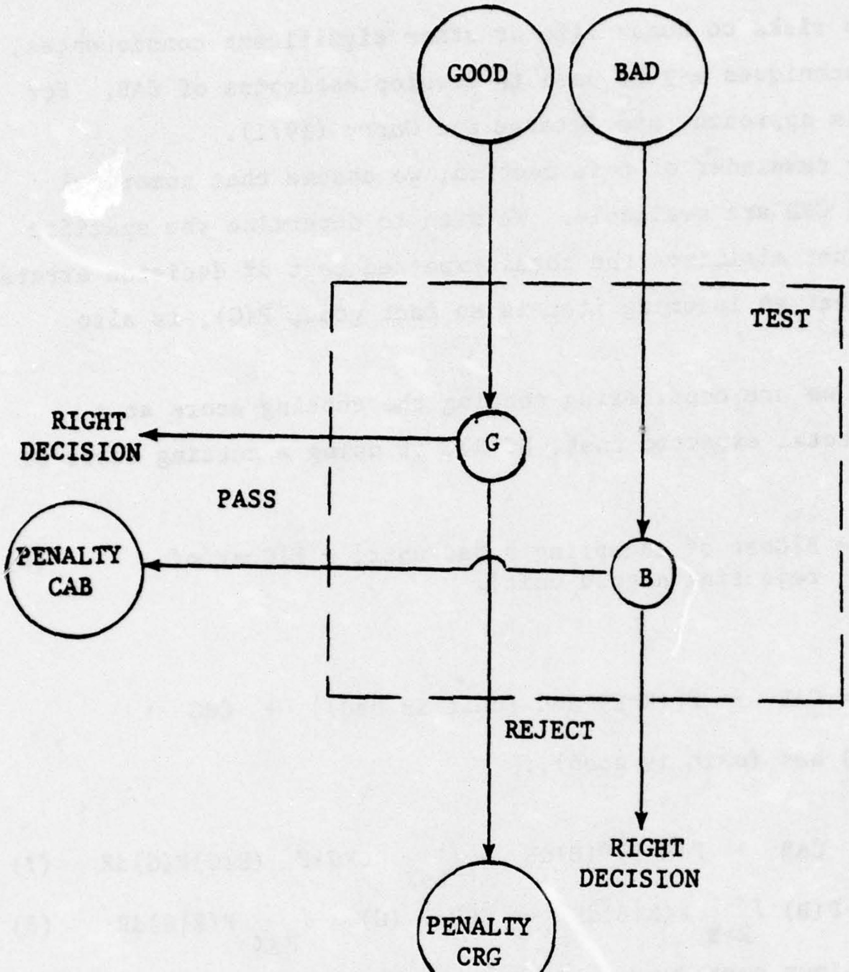


FIGURE 9. PENALTY COSTS FOR TEST DECISION ERRORS

bad item involves risks to human life or other significant consequences, utility theory techniques may be used to develop estimates of CAB. For an example of this approach, see Betaque and Gorry (1971).

For the remainder of this section, we assume that numerical values of CRG and CAB are available. We wish to determine the specific cutting score X that minimizes the total expected cost of decision errors. The probability that an incoming item is an fact good, P(G), is also assumed known.

Suppose we are considering setting the cutting score at a value of X. The total expected cost, EC(X), of using a cutting score of X is then

$$EC(X) = E[\text{Cost of accepting a bad unit}] + E[\text{Cost of rejecting a good unit}]. \tag{6}$$

Hence,

$$EC(X) = CAB \cdot P[(R>X) \text{ and } (\text{unit is bad})] + CRG \cdot P[(R<X) \text{ and } (\text{unit is good})].$$

$$= \int_{R>X} CAB \cdot P(R|B)P(B)dR + \int_{R<X} CRG \cdot P(R|G)P(G)dR \tag{7}$$

$$= CAB \cdot P(B) \int_{R>X} P(R|B)dR + CRG \cdot P(G) \cdot \int_{R<X} P(R|G)dR \tag{8}$$

The minimum cost cutting score occurs when the derivative of EC(X) with respect to X is zero. Hence,

$$\frac{dEC(X)}{dX} = 0 = -CAB \cdot P(B)P(X|B) + CRG \cdot P(X|G) = 0. \tag{9}$$

Hence, for the optimum cutting score,

$$P(X|B) = \frac{CRG \cdot P(G)}{CAB \cdot P(B)} P(X|G) \tag{10}$$

which may be written

$$P(X|B) = K \cdot P(X|G) \text{ where } K = \frac{CRG \cdot P(G)}{CAB \cdot P(B)} \tag{11}$$

Equation (11) simply says that the optimum cutting score X is such that the conditional probability of a reading X given the item is bad just equals a constant (K) times the conditional probability of a reading X given the item is good. The constant K is the ratio of the cost of rejecting all good items [CRG·P(G)] to the cost of accepting all bad items [CAB·P(B)].

Equation (11) provides a necessary condition for an optimal cutting score; note that this result is independent of the form of the probability distribution involved. To determine sufficient conditions for an optimal cutting score, however, the second derivatives of the cost function EC(X) must be investigated. In the next section, we present both necessary and sufficient conditions for the case of normally distributed error sources.

Normally Distributed Test Errors

Let us assume that the conditional distributions of $P(R|B)$ and $P(R|G)$ are normally distributed with means and variances of W_B, $\sigma_B{}^2$ and W_G, $\sigma_G{}^2$ for bad and good units, respectively; that is, $P(R|B) = \frac{1}{\sqrt{2\pi\sigma_B}}$ $\exp\left[-1/2 \left(\frac{X-W_B}{\sigma_B}\right)^2\right]$ and $P(R|G) = \frac{1}{\sqrt{2\pi\sigma_B}}$ $\exp\left[-1/2 \left(\frac{X-W_G}{\sigma_G}\right)^2\right.$

Equation (11) may then be written as

$$\frac{1}{\sqrt{2\pi\sigma B}} \quad \exp\left[1/2 \left(\frac{X-W_B}{\sigma_B}\right)^2\right] = K \quad \frac{\exp}{\sqrt{2\pi G}}\left[1/2 \left(\frac{X-W_G}{\sigma_G}\right)^2\right] \tag{12}$$

which may be rearranged into the form:

$$\exp\left[-1/2 \left(\frac{X-W_B}{\sigma_B}\right)^2\right] = K\cdot \frac{\sigma_B}{\sigma_G} \quad \exp\left[-1/2 \left(\frac{X-W_G}{\sigma_G}\right)^2\right] \tag{13}$$

Taking natural logarithms,

$$-1/2\left(\frac{X-W_B}{\sigma_B}\right)^2 = \ln \frac{K\sigma_B}{\sigma_G} - 1/2 \left(\frac{X-W_G}{\sigma_G}\right)^2 \tag{14}$$

Hence,

$$\frac{X^2-2W_B.X + W_B{}^2}{\sigma_B{}^2} = -2 \quad \ln\left(\frac{K\sigma_B}{\sigma_G}\right) + \frac{X^2-2W_GX + \sigma_G{}^2}{\sigma_G{}^2} \tag{15}$$

or

$$\left(\frac{1}{\sigma_G^2} - \frac{1}{\sigma_B^2}\right) X^2 - 2X\left(\frac{W_G}{\sigma_G^2} - \frac{W_B}{\sigma_B^2}\right) + \left(\frac{W_G^2}{\sigma_G^2} - \frac{W_B^2}{\sigma_B^2}\right)$$

$$-2\ln\left(\frac{K\sigma B}{\sigma G}\right) = 0 \tag{16}$$

Let

$$A = \left(\frac{1}{\sigma_G^2} - \frac{1}{\sigma_B^2}\right) \tag{17}$$

$$B = -2\left(\frac{W_G}{\sigma_G^2} - \frac{W_B}{\sigma_B^2}\right) \tag{18}$$

$$C = \frac{W_G^2}{\sigma_G^2} - \frac{W_B^2}{\sigma_G^2} - 2\ln\left(K\frac{\sigma_B}{\sigma_G}\right) \tag{19}$$

Using (17)-(19), equation (16) may be written as:

$$AX^2 + BX + C = 0 \tag{20}$$

From the quadratic formula, values of X which satisfy (20) are given by

$$X = \frac{-B \pm \sqrt{B^2-4AC}}{2A} \tag{21}$$

assuming $A \neq 0$.

Equation (21) identifies values of X for which the slope of the expected cost curve is zero. To determine if the associated cost is a local maximum or minimum, we may examine the second derivative. Specifically, since

$$\frac{dEC(X)}{d} = AX^2 + BX+C \tag{22}$$

then

$$\frac{d^2EC(X)}{d} = 2AX + B \tag{23}$$

From calculus, if X* satisfies (21), X* is a local minimum if (23) is negative. If (23) is positive, X* is a local maximum: finally, if (23) is zero, X* is an inflection point, and is neither a local maximum nor a local minimum.

Example 1

To illustrate the use of the above formulas, consider the situation illustrated to Figure 10. In this example, the failure variance is larger than the variance for good units, and 99% of the units being inspected are good. This would be the case, for example, for "on-aircraft" tests of items with low failure rates. For this example, assume the cost of accepting a bad unit (CAB) is $500, while the cost of rejecting a good unit (CRG) is $200.

Using the data from Figure 10

$$A = \left(\frac{1}{\sigma_G^2} - \frac{1}{\sigma_B^2}\right) = \left(\frac{1}{10^2} - \frac{1}{20^2}\right) = +.0075 \qquad (24)$$

$$B = -2\left(\frac{W_G}{\sigma_G^2} - \frac{W_B}{\sigma_B^2}\right) = -2\left(\frac{100}{10^2} - \frac{50}{20^2}\right) = -1.75 \qquad (25)$$

$$C = \frac{W_G^2}{\sigma_G^2} - \frac{W_B^2}{\sigma_B^2} - 2\ln\left(\frac{K\sigma_B}{\sigma_G}\right) = \frac{100^2}{10^2} - \frac{50^2}{20^2} - 2\ln$$

$$-2\ln\left(39.6\,\frac{20}{10}\right) = 85.01 \qquad (26)$$

since

$$K = \frac{CRG \cdot P(G)}{CAB \cdot P(B)} = \frac{200}{500} \cdot \frac{.99}{.01} = 39.6 \qquad (27)$$

Therefore, equation (16) yields

$$X = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} = \frac{1.75 \pm \sqrt{(1.75)^2 - 4(.0075)\,(88.67)}}{2(.0075)} =$$

$$68.9;\ 164.4 \qquad (28)$$

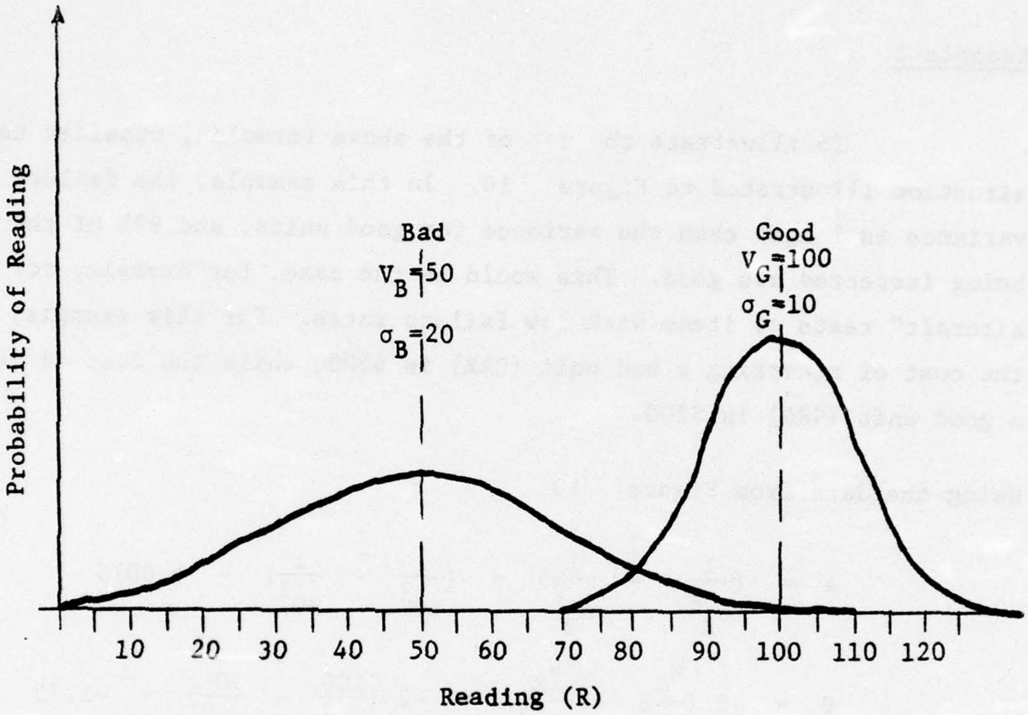


FIGURE 10. DISTRIBUTION OF TEST MEASUREMENTS FOR GOOD AND BAD COMPONENTS

The two candidates for the minimum cost cutting score are thus X=68.9 and X=164.4. When X = 68.9, 2AX+B = 2·(.0075)·(68.9) - 1.75 = -.716, a negative value.

Hence, X=68.9 is a local minimum. When X=164.4, 2AX-B=+.717, a positive value; hence, X=164.4 is a local maximum.

In this example, the expected cost EC(X) is:

$$EC(X) = CAB \cdot P(B) \int_{R>X} P(R|B)dR + CRG \cdot P(G) \int_{R<X} P(R|G)dR \tag{29}$$

$$= 5 \cdot \int_{R>X} P(R|B)dR + 198 \cdot \int_{R<X} P(R|G)dR \tag{30}$$

The expected cost for this example is plotted in Figure 11 and tabulated in Table 1. As can be seen from the figure, the minimum cost occurs at X=68.9; however, the expected cost is very close to the minimum for cutting scores in the range 60<X<75.

Example 2

Let us now suppose that 95% of the items coming to the test station are in fact bad. This would be the case faced by a field repair organization when fairly reliable Built in Test Equipment (BITE) is used to determine if a "squawked" item is to be removed from the aircraft and sent to field repair. Using the data from example 1 for the other

$$EC(X) = CAB \cdot P(B) \cdot \int_{R>X} P(R|B)dR + CRG \cdot P(G) \cdot \int_{R<X} P(R|G)dR \tag{31}$$

$$= 500 \cdot (.95) \cdot \int_{R>X} P(R|B)dR + 200 \cdot (.05) \cdot \int_{R<X} P(R|G)dR \tag{32}$$

$$= 475 \int_{R>X} P(R|B)dR + 10 \int_{R<X} P(R|G)dR \tag{33}$$

Specific values of EC(X) are tabulated in Table D-2, and plotted in Figure 12. By inspection of Figure 12, the minimum expected cost occurs with a cutting score of about X = 100. Let us see if this agrees with the analytical solution.

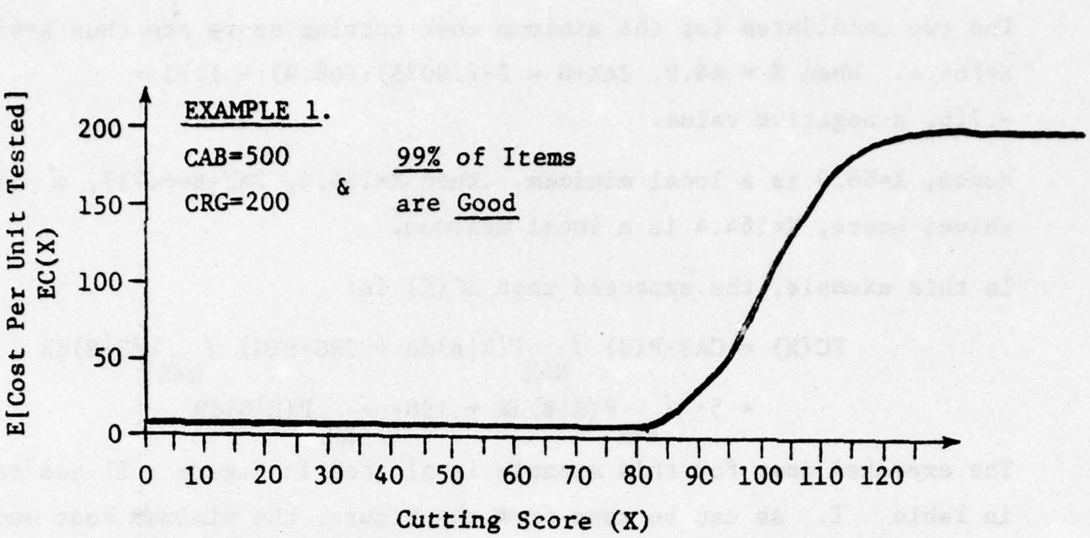


FIGURE 11. EXPECTED COST PER UNIT TESTED FOR EXAMPLE 1

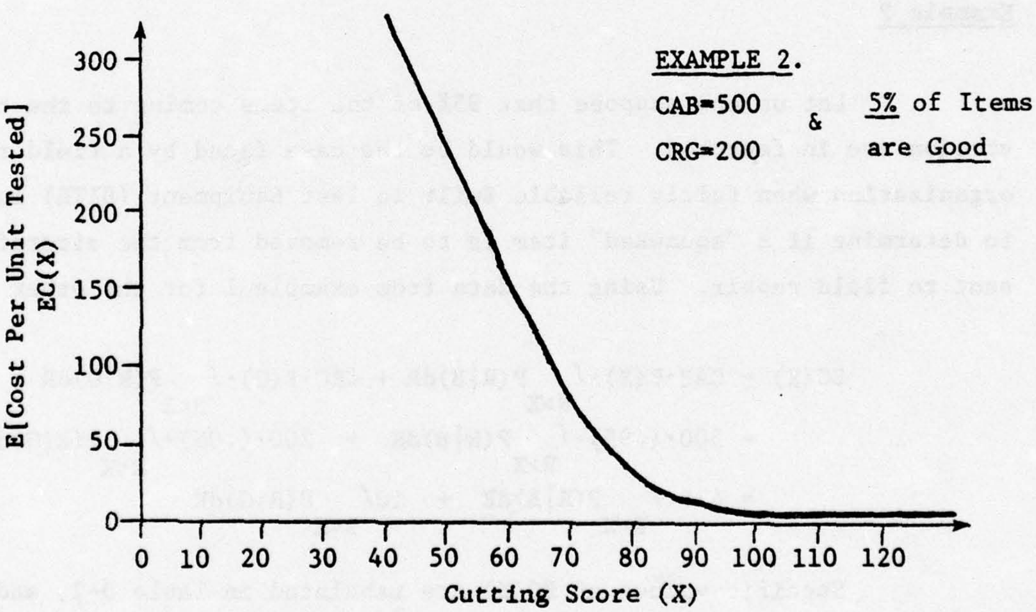


FIGURE 12. EXPECTED COST PER UNIT TESTED FOR EXAMPLE 2

Table 1. Expected Cost Calculations for Example 1

Column	1 X	2 $\frac{X-WB}{\sigma_B}$	3 R>XP(R|B)dR	4 5·3	5 $\frac{X-WG}{\sigma_G}$	6 R<XP(R|B)dR	7 198·6	8 4+7
	40	-0.5	.69	3.45	-6.0	0	--	3.4
	50	0.0	.50	2.5	- 5.0	0	--	2.5
	60	0.5	.31	1.5	-4.0	0	--	1.5
	68.9	.945	.1724	.862	-3.11	.0010	.198	1.06
	70	1.0	.1587	.793	-3.0	.0014	.271	1.07
	80	1.5	.07	.8	-2.0	.02	3.95	4.3
	90	2.0	.02	.1	-1.0	.16	31.68	31.7
	100	2.5	.01	.05	0	.50	99.00	99.0
	110	3.0	.001	.005	1.0	.84	166.32	166.3
	120	3.5	.0001	.005	2.0	.98	194.04	194.0
	140	4.5	0		4.0	1.00	198.	198.0
	160	5.5	0		6.0			
	170	6.0	0		7.0			

Table 2. Expected Cost Calculations for Example 2.

Column	1	2	3	4	5	6	7	8
	X	$\frac{X-W_B}{\sigma_O}$	$\int_{R>X} P(R \mid B)\,dR$	475·3	$\frac{X-W_G}{\sigma_G}$	$\int_{R<X} P(R \mid G)\,dR$	10·6	4+7
	30	-1.0	.84	399.	-7.0	0		399.
	40	-0.5	.69	327.	-6.0	--		327.
	50	0.0	.50	237.	-5.0	--		237.
	60	0.5	.31	147.	-4.0	--		147.
	70	1.0	.16	76.	-3.0	.0013	.013	76.
	80	1.5	.07	33.	-2.0	.023	.23	33.
	90	2.0	.023	10.9	-1.0	.16	1.6	12.5
	100	2.5	.0062	2.95	0	.50	5.0	8.0
	110	3.0	.0013	.62	1.0	.84	8.4	9.0
	120	3.5	.0001	.05	2.0	.98	9.8	9.9
	130	4.0	--	--	3.0	1.00	10.0	10.
	140	4.5	--	--	4.0	1.00	10.0	10.
	150	5.0	--	--	5.0	1.00	10.0	10.

Evaluating the parameters A, B, and C, we find

$$A = .0075 \qquad \text{(as in example 1)}$$
$$B = -1.75$$

and

$$K = \frac{CRG \cdot P(G)}{CAB \cdot P(B)} = \frac{200 \cdot (.05)}{500 \cdot (.95)} = .0210$$

Hence,

$$C = \frac{W_G 2}{\sigma_G^{\ 2}} - \frac{W_B 2}{\sigma_B^{\ 2}} - 2 \ln\left(K \frac{\sigma_B}{\sigma_G} \right) = 100 - 6.25 - 2 \ln (.042) = 100.09$$

Finally,

$$X = \frac{+1.75 \pm \sqrt{(1.75)^2 - 4 \cdot (.0075)\ (100.09)}}{2 \cdot (.0075)} = \frac{1.75 \pm .2445}{.015}$$

Hence,

$$X = 100.4;\ 133.0$$

Checking the second derivatives; $2AX + B = .244$ when $X = 100.4$; hence, $X = 100.4$ is a local minimum. When $X = 133.0$, $2AX + B = +.245$; thus $X = 133.0$ is a local maximum.

Special Case: $\underline{\sigma_G^{\ 2} = \sigma_B^{\ 2}}$

Consider the special case in which the item variances $\sigma_G^{\ 2}$ and $\sigma_B^{\ 2}$ are equal, i.e., $\sigma_G^{\ 2} = \sigma_B^{\ 2} = \sigma^{\ 2}$

In this case, coefficient (12) becomes zero. Hence, (15) may be written as

$$BX + C = 0$$

which gives

$$X = -C/B$$

$$= \frac{-\left[\frac{W_G^2 - W_B^2}{\sigma 2} - 2 \ln (K)\right]}{-2 \left(\frac{W_G - W_B}{\sigma 2}\right)}$$

$$= \frac{W_G^2 - W_B^2 - 2\sigma^2 \ln K}{2 \quad (W_G - W_B)}$$

$$= \frac{W_{G} + W_B}{2} - \frac{\sigma^2}{W_B - W_B} \ln K$$

Hence, when the item error variances are equal, the optimum cutting score is first set midway between the two means (the first term on the r.h.s) and then shifted in a direction determined by the factor K. If the cost of rejecting good items is greater than the cost of accepting bad items, ln K will be positive, and the optimum x will be less than the midpoint. Conversely, if the cost of accepting bad units is greater, ln K will be negative, and the optimum X is set above the midpoint.

For example, suppose

$W_G = 100$	CRG = 200	P(G) = .99
$W_B = 50$	CAB = 500	P(B) = .01
$\sigma = 10$		

Then $K = \ln \left(\frac{CRG \cdot P(G)}{CAB \cdot P(B)}\right) = \ln \frac{200 \cdot .99}{500 \cdot .01} = \ln (39.6) = 3.68$

The optimum cutting score X is then

$$X = \frac{100+50}{2} - \frac{2 \cdot (10)^2}{100-50} \cdot (3.68) = 75-15=60$$

Hence, for this situation, the optimum cutting score is 60, a value 15 points below the midpoint of the mean readings for good and bad items.

Parallel Test Facilities

The preceding analysis considered only single test facilities. In many instances, however, several test stations will work in parallel.

In this section, we ask "What is the relationship between cutting scores and the probability of passing the test for parallel test facilities?" As we shall see below, if we know the discriminating power of individual stations (as quantified by its C-Curve, as in Figure 8), then we can compute characteristic curves for the parallel facilities taken as a whole.

This will allow us, for analysis purposes, to treat a set of parallel stations as if it were a single facility. This means that the formulas developed in the previous section may be used to compute optimum cutting scores for both single and parallel facilities. In the remainder of this section, we show how to compute the C-Curve for parallel facilities when the C-Curve for the individual test stations are known.

Consider a test facility consisting of two parallel test stations, each with a known C-Curve. We assume that the probability that an incoming item is assigned to a specific station is independent of the specific condition of the item. This would be true, for example, if station assignments are based largely upon which station has the least amount of work in front of it.

Let us define the following events.

Event	Meaning
S_i	item is tested on station i
G	item is good
B	item is bad
P	item passes test
F	item fails test

Also, let P(X) denote the probability of X and let $P(X|E)$ denote the conditional probability of X given event E has occurred. Finally, let P(X&E) denote the joint probability of events X and E occurring together. First, let us compute the probability that a bad item passes the test. Let

$$P\,[\text{Bad item passes test}] = P[P|B] \tag{34}$$

But from elementary probability theory,

$$P[P|B] = \frac{P[B \;\&\; P]}{P(B)} \tag{35}$$

and

$$P[B\&P] = P[P\&B|S_1]P(S_1) + P[P\&B|S_2]P(S_2) \tag{36}$$

The conditional probabilities on the right may be computed since the C-Curves for each station are known, giving

$$P[P\&B|S_i] = \frac{P[P\&B\&S_i]}{P(S_i)} = \frac{P[P|B\&S_i]P[B\&S_i]}{P(S_i)} = P[P|B\&S_i]P(B) \tag{37}$$

since $P(B\&S_i) = P(B)\cdot P(S_i)$ assuming random station assignment. Re-writing (37), we have

$$P[P\&B|S_i] = P[P|B\&S_i]P(B) \tag{38}$$

Finally, substituting (38) into (36), and the result into (35), we have

$$P[P|B] = \frac{P[P|B\&S_1]P(B)\,P(S_1) + P[P|B\&S_2]P(B)P(S_2)}{P(B)} \tag{39}$$

$$= P[P|B\&S_1]P(S_1) + P[P|B\&S_2]P(S_2)$$

Equation (39) simply states that the probability of passing a bad item for the combination of stations may be obtained by weighting the probability that a given station passes a bad item by the probability that that station is used, and summing these products for all stations.

Note that if both stations have identical C-Curves, $P[P|B\&S_1] = P[P|B\&S_2]$, and the P[passing a bad item] for the combination of stations is equal to the probability of passing a bad item for an individual station.

These results are easily generalized. For N parallel stations, with random station assignment.

$$P[\text{Passing}|\text{Bad}] = \sum_{i=1}^{N} P[P|B\&S_i]P(S_i) \tag{40}$$

Similarly,

$$P[\text{Failing}|\text{Good}] = \sum_{i=1}^{N} P[F|G\&S_i]P(S_i) \tag{41}$$

V. MULTIVARIATE TESTS

To this point we have assumed that the result of a given test was summarized in a single reading R. Let us now consider more complex tests in which several physical or electrical measurements of a given item are obtained. Based on these measurements, we are to either accept or reject the item. As before, we assume that the item is in fact either good or bad, but that the true state or condition of the item cannot be directly observed. The question is, "How can optimum cutting scores be established in this case?"

Let r_i denote the value of the ith measurement or reading for a particular item, and let R denote the vector of measurements obtained during a given test. That is,

$$R = (r_1, r_2, \ldots, r_N)$$

where N denotes the number of measurements taken. Let CAB denote the cost of accepting a bad item, and let CRG represent the cost of rejecting a good item. Then the following decision table describes the current situation.

Decision Table Given a Reading Vector R

Item is in fact	Probability	Possible Actions: Accept Item	Possible Actions: Reject Item
Good	$P(G\|R)$	0	CRG
Bad	$P(B\|R)$	CAB	0

In the above table, $P(G|R)$ denotes the conditional probability that the item is good given the reading vector R is observed, while $P(B|R)$ denotes the conditional probability the item is bad.

Using the Bayes decision criteria, we would accept the item if the expected loss due to accepting bad items is less than the expected loss due to rejecting good items. That is,

accept if

$$CAB \cdot P(B|R) < CRG \cdot P(G|R) \tag{42}$$

But the probability the item is bad equals one minus the probability the item is good, i.e.

$$P(B|R) = 1 - P(G|R) \tag{43}$$

Therefore, we would accept if

$$CAB \cdot [1 - P(G|R)] < CRG \cdot P(G|R) \tag{44}$$

Rearranging, we obtain

$$P(G|R) > \frac{CAB}{CAB+CRG} \tag{45}$$

Hence, we should accept the item if the conditional probability the item is good exceeds the ratio of CAB to [CAB + CRG]. Conversely, if $P(G|R)$ is less than this ratio, we would reject the item. Finally, if $P(G|R)$ equals this ratio, we would be indifferent between accepting or rejecting the item, since the expected cost of each decision is the same.

The conditional probability $P(G|R)$ for a given vector of readings R may be computed using Bayes theorem. Specifically,

$$P(G|R) = \frac{P(R|G)P(G)}{P(R|G)P(G)+P(R|B)P(B)} \tag{46}$$

where $P(R|G)$ and $P(R|B)$ denote the probability distributions of readings from good and bad items, respectively, and $P(G)$ and $P(B)$ denote the respective <u>a priori</u> probabilities that an item is in fact good or bad.

Suppose X is a set of readings such that

$$P(G|X) = \frac{CAB}{CAB+CRG} \tag{47}$$

For notational convenience, let

$$Z = P(G|X). \tag{48}$$

Then from Bayes theorem,

$$Z = \frac{P(X|G)P(G)}{P(X|G)P(G)+P(X|B)P(B)} \tag{49}$$

Rearranging,

$$Z \cdot P(X|G)P(G)+Z \cdot P(X|B)P(B) = P(X|G)P(G) \tag{50}$$

$$Z \cdot P(X|B)P(B) = (1-Z) \cdot P(X|G)P(G) \tag{51}$$

Hence,

$$P(X|B) = \frac{(1-Z)}{Z} \cdot \frac{P(G)}{P(B)} \cdot P(X|G) \tag{52}$$

But recall that $Z=P(X|G) = CAB|[CAB+CRG]$. Hence, it can be shown that

$$\frac{(1-Z)}{Z} = \frac{CRG}{CAB} \tag{53}$$

Finally, substituting (53) into (52),

$$P(X|B) = \frac{CRG \cdot P(G)}{CAB \cdot P(B)} \cdot P(X|G) \tag{54}$$

which may be written as

$$P(X|B) = K \cdot P(X|G) \tag{55}$$

where

$$K = \frac{CRG \cdot P(G)}{CAB \cdot P(B)} \tag{56}$$

The above result is the multivariant equivalent of the necessary condition for optimum cutting scores given by equation (11) in Section IV.

In both the single and multivariate case, we seek values of X such that the conditional probability of observing X for a bad item equals a constant K times the conditional probability of observing X if the item is in fact good. The constant K equals the ratio of the cost of rejecting all good items [CRG·P(G)] to the cost of accepting all bad ones [CAB·P(B)].

In the single measurement case, relation (55) will yield a unique value for X for most probability distributions encountered in practice. In the multivariate case, however, we may expect a large number of sets of readings $X = (X_1, X_2, \ldots, X_N)$ to satisfy the above condition.

For example, consider a test in which two readings, r_1 and r_2, are taken. Furthermore, suppose the readings are statistically independent. Then

$$P(R|B) = P(r_1|B) \cdot p(r_2|B) \tag{57}$$

and

$$P(R|G) = P(r_1|G) \cdot P(r_2|G) \tag{58}$$

Let X_1 and X_2 denote the cutting scores to be used for comparison with readings r_1 and r_2, respectively. Then from (55), X_1 and X_2 must satisfy

$$P(X_1|B)\, P(X_2|B) = K \cdot P(X_1|G) \cdot P(X_2|G) \tag{59}$$

If a specific value for X_1 is selected, then X_2 must satisfy

$$P(X_2|B) = \frac{K \cdot P(X_1|G)}{P(X_1|B)} \cdot P(X_2|G) \tag{60}$$

Thus, even in this simple case, the choice of the optimum cutting score X_2 for the second reading is dependent upon the cutting score X_1 used for the first reading.

In general, specification of optimum cutting scores for a multivariate test requires identification of all sets of values $X = (X_1, X_2, \ldots, X_N)$ satisfying equation (55). In the two-reading case, this may be done graphically, as illustrated in Figure 13. Any item with a combination of readings, lying inside the "Acceptable Region" would pass the test, and items with readings outside this region would fail. In this example, "rough" cutting scores may be established using the extreme measurements in the Acceptable Region.

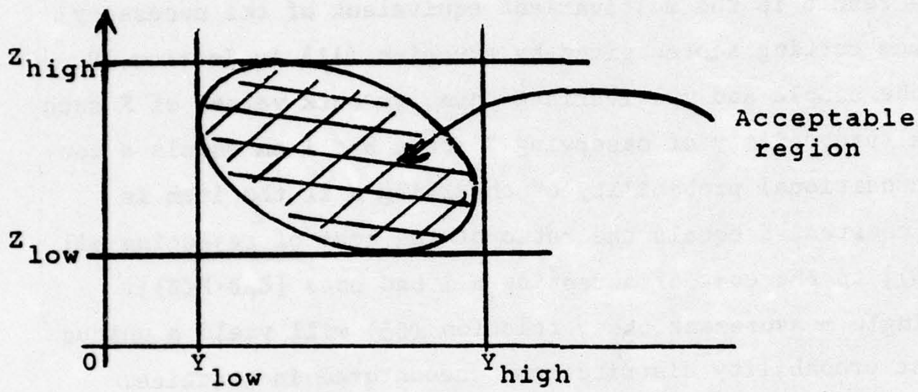


FIGURE 13. THE ACCEPTABLE REGION FOR A TWO-READING TEST

For example, in Figure 13 if the measurement Y is taken first, the values marked Y_{low} and Y_{high} might be used to determine if measurement Z should be taken. If the measurement Y is less than Y_{low}, more than Y_{high}, the item will fail the test, regardless of the value of Z obtained. Hence, if Y is outside the range $[Y_{low}, Y_{high}]$, the item should be rejected immediately. On the other hand, if the reading is within this range, Z must also be measured to determine if the pair of readings (Y,Z) is in the acceptable region.

When more than two readings are involved, graphical methods can no longer be used to specify the acceptable region. In this case, two major options are available.

1. Explicitly specify all reading vectors X in the acceptable range, using tables or monographs
2. Determine whether to accept or reject an item by first obtaining the reading vector R, and then computing $P(G|R)$ directly.

To illustrate the second approach, suppose CAB = \$500 and CRG = \$200. From equation (45), after obtaining a given reading vector R, we should accept the item if

$$P(G|R) = \frac{CAB}{CAB+CRG} = \frac{500}{500+200} = .74 \tag{61}$$

Hence, if the conditional probability $P(G|R) > .74$, we should accept the item; otherwise, we should reject it.

To illustrate the computation of $P(G|R)$, suppose we take two measurements r_1 and r_2 that are bivariate normally distributed. Specifically, suppose that if an item is in fact good,

$$P(R|G) = K_G \cdot \exp\left[-1/2(R-U_G)'S_G\,(R-U_G)\right] \tag{62}$$

where

$$R = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} ; \; U_G = \begin{pmatrix} 100 \\ 100 \end{pmatrix} ; \; V_G = \begin{bmatrix} 100 & 25 \\ 25 & 100 \end{bmatrix} \tag{63}$$

where R denotes the reading vector, U_G denotes the vector of expected readings for good items, V_G denotes the variance - covariance matrix describing the joint variability of readings r_1 and r_2, and $S_G = V_G^{-1}$ by definition. (For a detailed discussion of the properties of multivariate normal distributions, see F.A. Graybill, An Introduction to Linear Statistical Models, Volume I, McGraw-Hill, 1961). The value of the integration K_G is given by

$$K_G = \frac{\pi_{i=1}^{N} \lambda_i^{1/2}}{(2\pi)\; N/2} \tag{64}$$

where λ_i is the ith characteristic root of S_G, and N denotes the number of readings. In our case, N = 2.

For this example,

$$S_G = V_G^{-1} = \begin{bmatrix} .01067 & .00267 \\ .00267 & .01067 \end{bmatrix} \tag{65}$$

By definition, the characteristic root λ of S_G satisfies, for any vector X,

$$S_G \cdot X = \lambda X \tag{66}$$

which implies

$$|S_G - I| = 0 \tag{67}$$

Hence, for S_G,

$$|S_G - I| = \begin{vmatrix} .01067-\lambda & .00267 \\ .00267 & .01067-\lambda \end{vmatrix} \quad (68)$$

$$= (.01067-\lambda)^2 - (.00267)^2 = 0 \quad (69)$$

therefore,

$$.01067-\lambda = \pm .00267 \quad (70)$$

which gives

$$\lambda = .00800 \text{ or } .01334 \quad (71)$$

Hence,

$$K_G = \frac{\pi_{i=1}^{2} \lambda_i^{1/2}}{2\pi} = \frac{(.08944)\ (.115499)}{2\pi} = .0016441 \quad (72)$$

Hence, given any set of readings r_1, and r_2, we may evaluate $P(R|G)$ by using the values defined by equations (63), (65), and (72) in equation (63).

Similarly, <u>if the item is in fact bad</u>, let us assume that the reading vector R is also bivariate normally distributed, with the following parameters:

$$P(R|B) = K_B \exp\ [-1/2(R-U_B)'S_B\ (R-U_B)] \quad (73)$$

where

$$U_B = \begin{pmatrix} 50 \\ 50 \end{pmatrix};\ V_B = \begin{pmatrix} 225 & 100 \\ 100 & 225 \end{pmatrix};\ S_B = V_B^{-1} = \begin{bmatrix} .005538 & .00246 \\ .00246 & .005538 \end{bmatrix} \quad (74)$$

In this case, the characteristic roots of S_B satisfy

$$\begin{vmatrix} .005538-\lambda & .00246 \\ .00246 & .005538-\lambda \end{vmatrix} = 0 \quad (75)$$

Hence,

$$(.005538-\lambda)^2 - (.00246)^2 = 0 \quad (76)$$

or

$$.005538-\lambda = \pm .00246 \quad (77)$$

Finally,

$$= .00800 \text{ or } .00308 \quad (78)$$

Hence, from (64),

$$K_B = \frac{(.00800)^{1/2}\ (.00308)^{1/2}}{2\Pi} \quad \frac{(.089442)\ (.05548)}{2\Pi} = .000790 \quad (79)$$

Finally, suppose that we obtain readings of $r_1=85$ and $r_2=70$, and suppose that the <u>a priori</u> probability that the item is good is $P(G) = .99$. Then, evaluating $P(R|G)$ using equations (62), (63), and (72), we obtain

$$R-U_G = \begin{pmatrix} 85-100 \\ 70-100 \end{pmatrix} = \begin{pmatrix} -15 \\ -30 \end{pmatrix}$$

$$(R-U_G)'S_G(R-U_G) = (-15, -30)\begin{pmatrix} .01067 & .00267 \\ .00267 & .01067 \end{pmatrix}\begin{pmatrix} -15 \\ -30 \end{pmatrix}$$

$$= (-.24015, -.48015)\begin{pmatrix} -15 \\ -30 \end{pmatrix} = 18.006$$

Therefore,

$$P(R|G) = .0016441 \exp[-1/2(18.006)]$$
$$= (.0016441)\ (.000123)$$
$$= .000\ 000\ 20222 = 2.0222 \times 10^{-7}$$

Also, evaluating $P(R|B)$ using (73), (74), and (79), we obtain

$$\begin{pmatrix} 85-50 \\ 75-50 \end{pmatrix} \qquad \begin{pmatrix} 35 \\ 25 \end{pmatrix}$$

$$(R-U_B)'S_G(R-U_B) = (35,25)\begin{pmatrix} .005538 \\ .00246 \end{pmatrix}\begin{pmatrix} .00246 \\ .005538 \end{pmatrix}\begin{pmatrix} 35 \\ 25 \end{pmatrix}$$

$$= (.255225, .22455)\begin{pmatrix} 35 \\ 25 \end{pmatrix}$$

$$= 14.547$$

Hence,

$$P(R|B) = (.000790) \exp[-1/2\ (14.547)]$$
$$= (.000790)\ (.000694)$$
$$= .000\ 000\ 5482 = 5.482 \times 10^{-7}$$

Finally,

$$P(G|R) = \frac{P(R|G)P(G)}{P(R|G)P(G)+P(R|B)P(B)}$$

$$= \frac{(.000\ 000\ 2022)\ (.99)}{(.000\ 000\ 2022)\ \ (.99) + (.000\ 000\ 5482)\ (.01)}$$

$$= \frac{2001.78}{2001.78 + 54.82} = .973$$

Hence, with readings of 85 and 70, there is a 97% chance that the item is good. Note that our a priori probability that the item was good was 99%; hence, we have revised the probability that this particular item is good down a little based on the readings. However, 97% is still much more than our 74% acceptance limit; hence, this item would be accepted.

Now suppose a second item is tested. In this case, suppose the readings are $r_1 = 60$ and $r_2 = 50$. In this case,

$$R-U_G = \begin{pmatrix} 60-100 \\ 50-100 \end{pmatrix} = \begin{pmatrix} -40 \\ -50 \end{pmatrix}$$

$$(R-U_B)'S_G(R-U_G) = (-40, -50) \begin{pmatrix} .01067 & .00267 \\ .00267 & .01067 \end{pmatrix} \begin{pmatrix} -40 \\ -50 \end{pmatrix}$$

$$= (-.5603, -.6403) \begin{pmatrix} -40 \\ -50 \end{pmatrix}$$

$$= 54.427$$

Hence,

$$P(R|G) = .001644 \exp [-1/2(54.427)]$$

$$= (.001644)\ (1.512 \times 10^{-12})$$

$$= 2.486 \times 10^{-15}$$

Next, let us compute $P(R|B)$. We obtain

$$R-U_B = \begin{pmatrix} 60-50 \\ 50-50 \end{pmatrix} = \begin{pmatrix} 10 \\ 0 \end{pmatrix}$$

$$(R-U_B)'S_B(R-U_B) = (10,0) \begin{pmatrix} .005538 & .00246 \\ .00246 & .005538 \end{pmatrix} \begin{pmatrix} 10 \\ 0 \end{pmatrix}$$

$$= (.05538, .0246) \begin{pmatrix} 10 \\ 0 \end{pmatrix}$$

$$= .5538$$

Hence,

$$P(R|B) = .000790 \exp [-1/2(.5538)]$$

$$= (.000790)\ (.758131)$$

$$= .005989$$

Finally,

$$P(G|R) = \frac{(2.48 \times 10^{-15})\ (.99)}{(2.48 \times 10^{-15})\ (.99) + (.005989)\ (.01)}$$

$$= 4.10 \times 10^{-9} \doteq 0$$

Hence, the second item is almost certainly bad, and should be rejected.

VI. NETWORKS OF TEST AND REPAIR FACILITIES

The previous analysis was concerned with setting optimum cutting scores for a single test. For many complex items, however, cost-effective logistics support requires development of diagnostic and repair facilities at several physical locations. In this section, we discuss an approach for determining optimal cutting scores for networks of test and repair facilities.

A FIELD-TO-DEPOT FLOW MODEL

To begin, let us consider an airborne item supported by a logistics network such as that shown in Figure 14. For this item, tests are performed in the field after each mission is flown. If no problems are discovered, the unit remains on the aircraft. If the tests indicate a problem, however, the unit is removed from the aircraft and shipped to the depot for further testing and repair. After all known problems are resolved, the unit is returned to the field and the cycle continues.

Our current objective is to estimate the cost of supporting 1,000 sorties if a given combination of cutting scores and test and repair facilities are used. Later, we will discuss how to determine optimal cutting scores for a network of facilities.

To simplify our discussion, suppose that for our hypothetical item the same tests and procedures are used at both the field and depot level. Suppose these tests have the following characteristics:

Value	Meaning
50	Average reading for failed items
100	Average reading for good items
17	Standard deviation of readings.

Furthermore, suppose a cutting score of 75 is used at both field and depot levels. Assuming the readings are normally distributed,

$$P\ [\text{accepting a bad item}] = P\left[Z \leq \frac{50-75}{17}\right] = .07$$

$$P\ [\text{rejecting a good item}] = P\left[Z \leq \frac{100-75}{17}\right] = .07$$

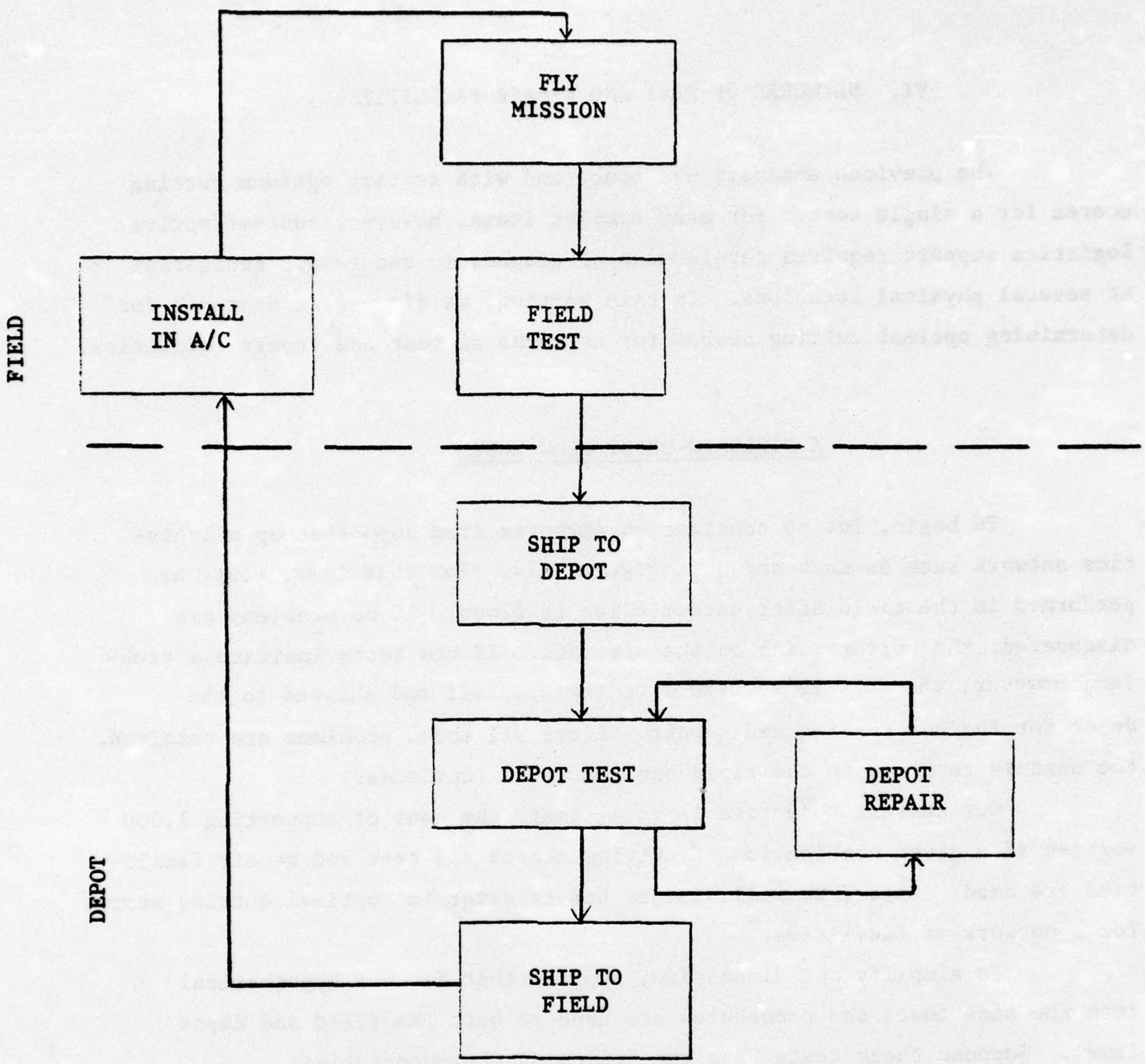


FIGURE 14. A TWO-LEVEL TEST AND REPAIR NETWORK

Hence, for this example, a cutting score of 75 results in error probabilities of 7%, at both field and depot levels, while the chance of a correct decision for any given item is 93%.

To continue our example, suppose that the repair process is not perfect; specifically, let us assume that there is a 5% chance that an item that is sent to repair will be bad at the conclusion of repair actions; conversely, there is a 95% chance that repair is successful. If a good item is sent to repair (due to testing error), we assume the same probabilities apply.

Finally, suppose that the item has a mean time between failures of 200 hours, with a constant hazard rate; this implies that the probability the item is used for at least X hours before failure is given by

$$P[\text{no failure in X hours}] = e^{\frac{-X}{200}}$$

If the average sortie length is two hours, this implies that the probability of no failure on any specific mission is $e^{-.01} = .99$; hence, the probability of failure is .01.

With these assumptions, the possible conditions of the unit may be described by the state diagram shown in Figure 15. In this diagram, circles denote particular states of the unit while arrows denote possible changes of state, or transitions, that the unit may follow. The "G" indicates states in which the component is in fact good, while "B" indicates states in which the component is in fact bad. Note that the diagram has been constructed such that all "good" states are on the left, and all "bad" states are on the right.

The following symbols denote the particular physical location associated with a given item state.

Symbol	Physical Location
S	Installed in aircraft; on a mission
FT	Field Test
SD	Ship to Depot; item is in transit
DT	Depot test
DR	Depot repair
SF	Ship to field

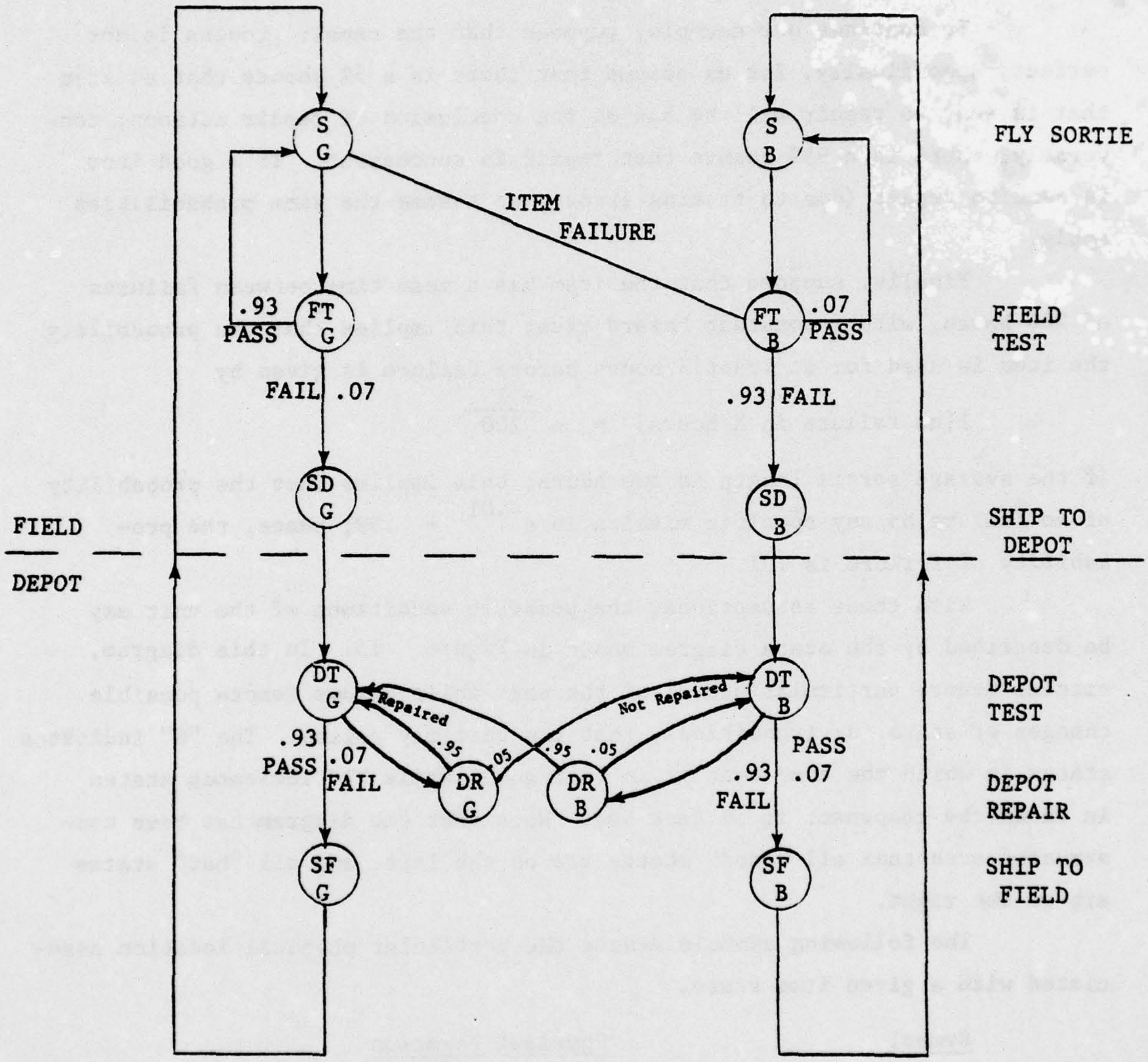


FIGURE -15. STATE DIAGRAM FOR A SIMPLE FIELD-TO-DEPOT FLOW MODEL

Hence, the circle marked "S,G" denotes the state in which the item is good, and is flying a mission. Similarly, the state "FT,G" indicates the item is being tested in the field following the sortie, and is still good.

In Figure 15, the numbers beside each arrow denote the probability that a given path is followed when the item changes state. Hence, the .01 by the arrow connecting states "S,G" and "FT,B" indicates there is a 1% chance that the item will fail on any given sortie, and hence bad ("B") when it enters the post-sortie field test ("FT"). Similarly, the .93 beside the arrow connecting state ("FT,G") to state "S,G" indicates there is a 93% chance that a good item will pass the field test ("FT"); while, conversely, there is a 7% chance that a good item will mistakenly fail the test and be shipped to the depot (state "SD,G").

Evaluating Expected Costs

This flow model may be used to determine the expected cost of maintaining a given number of operational units with a given combination of field and depot cutting scores. Jointly optimal cutting scores may then be determined by evaluating all possible combinations of cutting scores for the field and depot tests. Optimization procedures might also be used to intelligently select specific cutting score values for evaluation. We will discuss one such optimization approach later in this section.

Let us now consider how the expected cost of a specific configuration of cutting scores may be determined. Suppose we assign numbers 1, 2, ..., N to each state represented by a circle in Figure 15. Suppose that:

C_j = cost of passing through state j

π_j = expected number of times a unit passes through state j

P_{ij} = probability that a component that is in state i transitions to state j.

Then, in steady state,

$$\Pi_j = \sum_{i=1}^{N} \Pi_i P_{ij} \qquad \text{for } j = 1, 2, \ldots, N \tag{80}$$

That is, when the system is in steady state, the expected number of times that a unit passes through state j (i.e. Π_j) equals the expected number of times the unit passes through state i, (Π_i) multiplied by the the probability, (P_{ij}) that a unit next transitions from state i to state j. The products of Π_i P_{ij} are then summed over all possible prior states.

Suppose we number the states shown in Figure 15, beginning with state "S,G", and continuing until all states are identified. Hence, Π_1 denotes the number of successful missions flown. If the system of equations (80) is solved with the additional requirement that

$$\Pi_1 = 1000 \tag{81}$$

the resulting values of Π_j denote the expected number of times that a given unit will pass through state j before 1000 successful missions are completed.

For example, the solution to the system of equations (80) and (81) using the data from Figure 15 is shown in Figure 16. The numbers written beside each state j denotes the corresponding value of Π_j. For example, notice that for this particular configuration of tests and repair facilities, for every 1000 successful missions flown there will be 69 good units that are mistakenly shipped to the depot for repair, and an average of 10.8 bad units shipped to the depot. Since the depot test is also imperfect, we would expect 6 units that are in fact good to be mistakenly sent to repair. In contrast, the expected number of bad units sent to repair is 10.8 per 1000 successful missions. Hence, for this system of tests, 6 (6 + 10.8), or 36% of all repair activity is on <u>good</u> items.

If the cost C_j of having an item pass through each state j is known, the cost of supporting 1000 successful missions, C, is simply

$$C = \sum_j C_j \Pi_j \tag{82}$$

For example, suppose the following cost calculations apply to Figure 16.

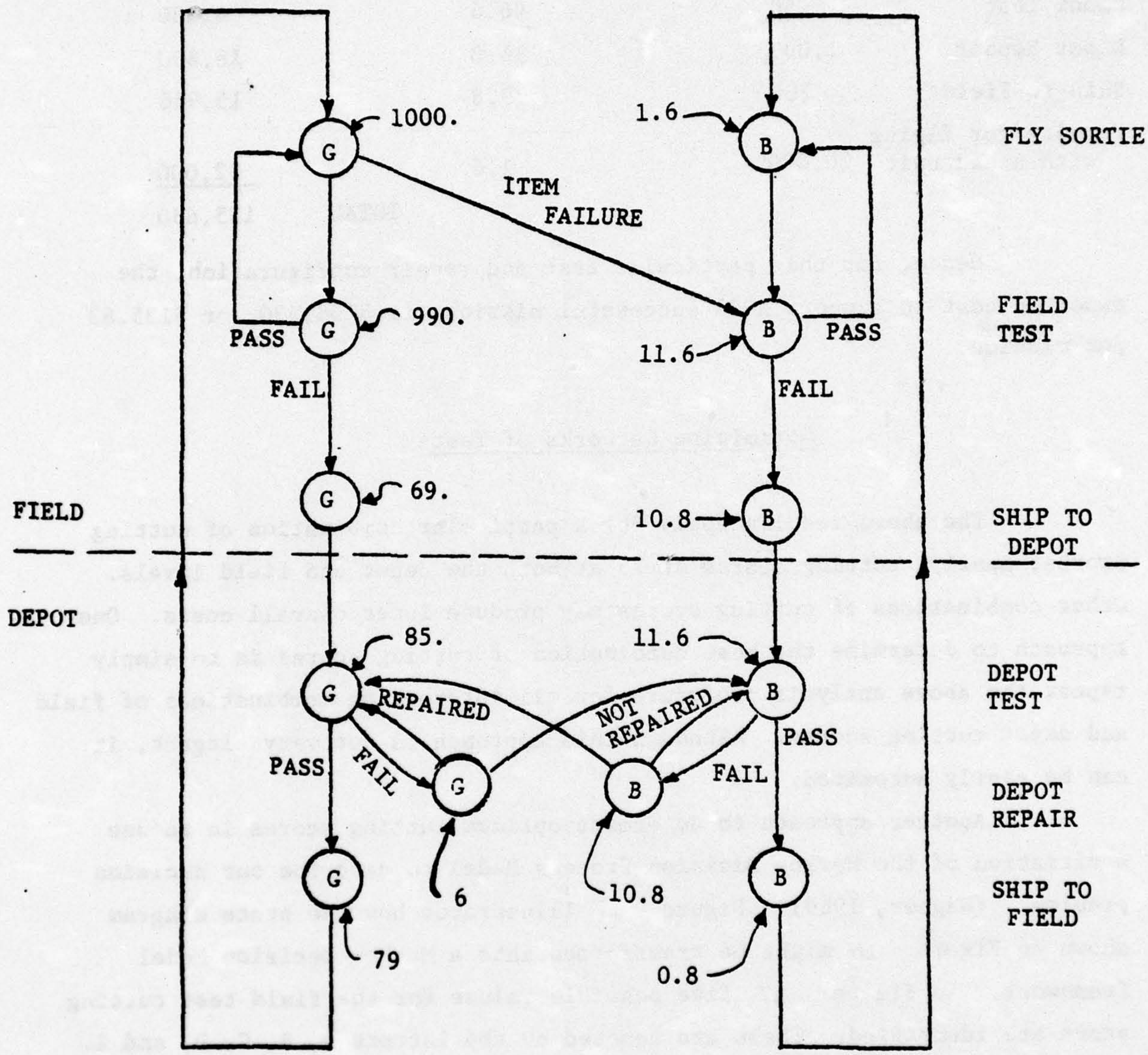


FIGURE 16. FLOWS PER 1000 SUCCESSFUL SORTIES

Action	Cost	Frequency Per 1000 Missions	Cost x Frequency
Field Test	50	1001.6	50,080
Ship to Depot	200	79.8	15,960
Depot Test	50	96.6	4,830
Depot Repair	1,000	16.8	16,800
Ship to Field	200	79.8	15,960
Penalty for flying with a bad unit	20,000	1.6	32,000
		TOTAL	135,630

Hence, for this particular test and repair configuration, the expected cost to support 1000 successful missions is \$135,630, or \$135.63 per mission.

Optimizing Networks of Tests

The above results apply for a particular combination of cutting scores; namely, *cutting scores of 75 at both the depot and field levels.* Other combinations of cutting scores may produce lower overall costs. One approach to determine the best combination of cutting scores is to simply repeat the above analysis procedure for all interesting combinations of field and depot cutting scores. Although this approach is not very elegant, it can be easily automated.

Another approach to determine optimum cutting scores is to use a variation of the Markov Decision Process Model to describe our decision problem. (Wagner, 1969). Figure 17 illustrates how the state diagram shown in Figure 16 might be transformed into a Markov Decision Model framework. In Figure 17, five possible values for the field test cutting score are identified. These are denoted by the letters A, B, C, D, and E. Similarly, five possible depot cutting scores (I, J, K, L, and M) are identified.

Suppose we number each state shown in Figure 17, starting with state "S,G". Let C_j, Π_j, and P_{ij} be defined as above. Then the optimum set of cutting scores may be determined by solving the following mixed integer programming problem for the values of Π_j and Y_i that minimize C, that is:

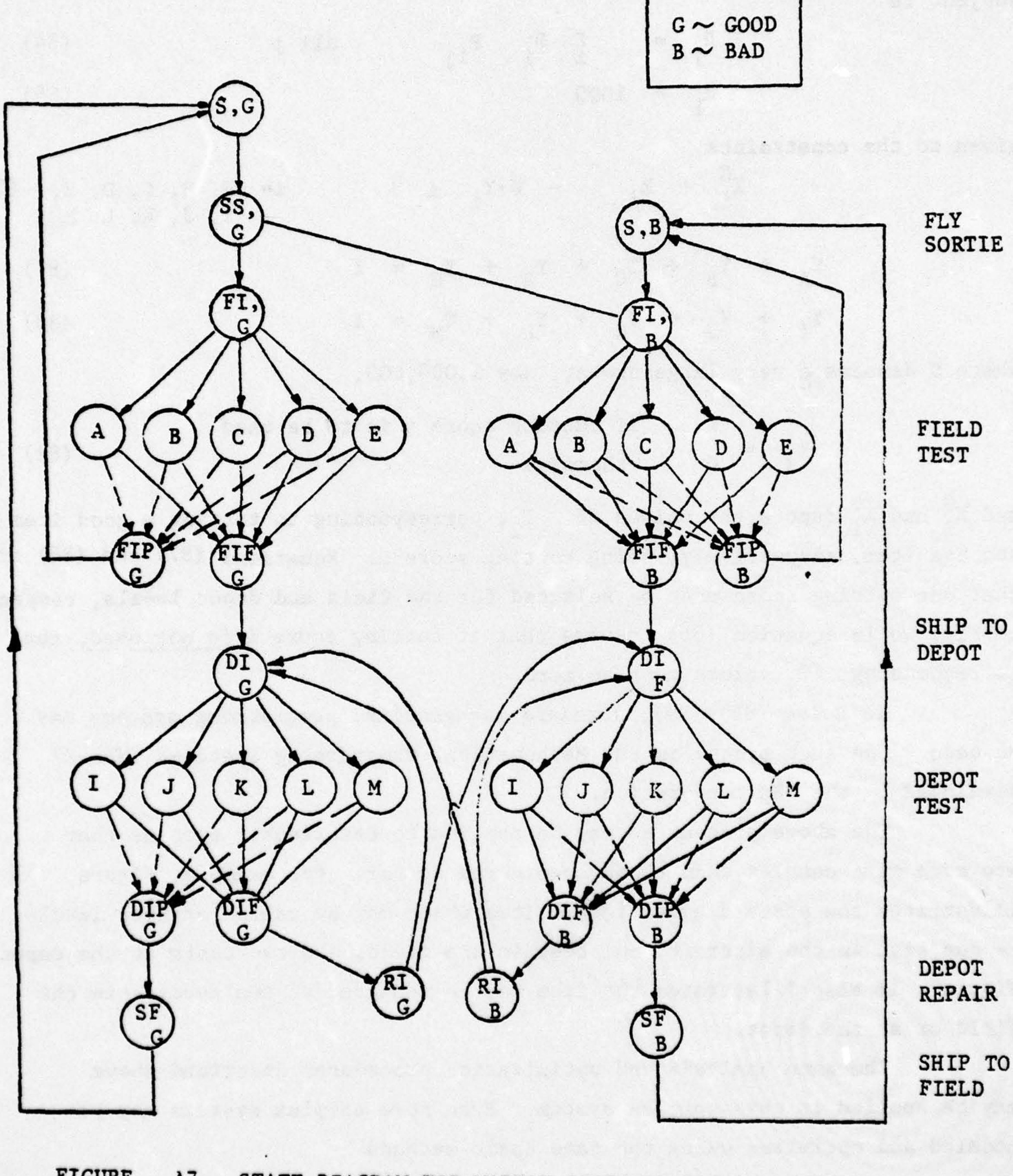


FIGURE 17. STATE DIAGRAM FOR MARKOV DECISION PROCESS MODEL

$$C = \sum_j C_j \Pi_j \tag{83}$$

subject to

$$\Pi_j = \sum_i \Pi_i P_{ij} \qquad \text{all } j \tag{84}$$

$$\Pi_1 = 1000 \tag{85}$$

given to the constraints

$$X_i^G + X_i - U \cdot Y_i \leq 0 \qquad i = A, B, C, D, E, I, J, K, L, M \tag{86}$$

$$Y_A + Y_B + Y_C + Y_D + Y_E = 1 \tag{87}$$

$$Y_I + Y_J + Y_K + Y_L + Y_M = 1 \tag{88}$$

where U denotes a very large number, say 1,000,000,

$$Y_i = \begin{cases} 1 & \text{if cutting score } i \text{ is to be used} \\ 0 & \text{otherwise} \end{cases} \tag{89}$$

and X_i^G and X_i^B denote the values of Π_2, corresponding to testing a good item and bad item, respectively, using cutting score i. Equations (87) and (88) state that one cutting score must be selected for the field and depot levels, respectively, while equation (86) insures that if cutting score i is not used, the corresponding Π_j values must be zero.

To solve (83)-(89), standard mathematical programming systems may be used. One such system is the Mathematical Programming System-X (MPS-X) developed by the IBM Corporation.

The above procedures may be applied to test/repair systems that are much more complex than those considered so far. For example, Figure 18 illustrates the state diagram for an item which may be tested at four levels -- one test in the aircraft, one test in the field, and two tests at the depot. Figure 18 also illustrates the item can be repaired at two levels--in the field or at the depot.

The same analysis and optimization procedures described above may be applied to this complex system. Even more complex systems may be modeled and optimized using the same basic methods.

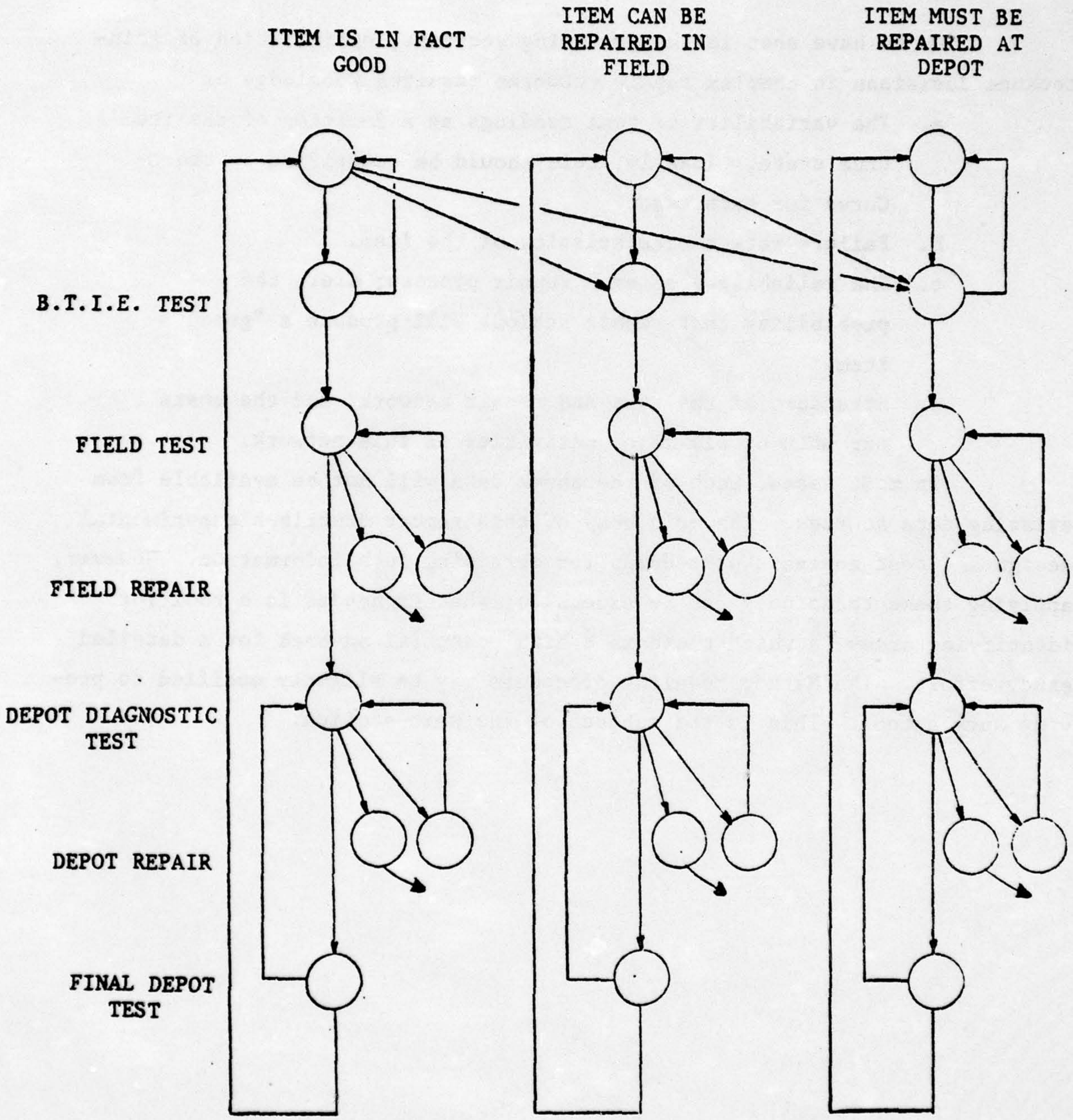


FIGURE 18. STATE DIAGRAM FOR MULTI-LEVEL TEST AND REPAIR

AD-A032 205 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OHIO F/G 12/2
PROCEEDINGS OF THE SYMPOSIUM ON APPLICATION OF DECISION THEORY --ETC(U)
OCT 76 N K WOMER
UNCLASSIFIED AFIT-TR-76-15 NL

3 OF 3
AD
A032205

END
DATE
FILMED
1-77

Data Collection

As we have seen in the preceding sections, optimization of maintenance decisions in complex repair networks requires knowledge of

a. The variability of test readings as a function of the item's true state. Ideally, this should be quantified by the C-Curve for each test.

b. Failure rate characteristics of the item.

c. The reliability of each repair process; i.e., the probability that repair actions will produce a "good" item.

d. Structure of the test and repair network, and the costs per unit of all major activities in this network.

In most cases, much of the above data will not be available from existing data sources. The main body of this report describes experimental design and cost analysis procedures for obtaining such information. However, applying these techniques can be expensive; what is needed is a tool for identifying areas in which there is a high potential payback for a detailed study effort. The Markov Modeling procedure may be slightly modified to provide such a tool. This is the subject of the next section.

VII. PROCESS FLOW MODELS

Repair systems for the KT-73 and for many other Air Force items are extremely complex; involving many specific repair activities and diagnostic and performance tests. A fundamental problem in improving the effectiveness of these complex systems is identification of areas with high payback potential, that is

> "...[to determine] the general areas of subprocesses which would, if improved slightly, result in major improvements in cost effectiveness, and conversely, [to determine] areas or subprocesses which, if greatly improved, would have little effect on overall cost effectiveness" (Genet, July, 1970).

In most complex systems, a very few areas are the source of the majority of system problems and system costs. Usually, the most effective approach to improve the performance of a complex system is by concentrating attention on these "critical few". The question is, "How can we identify these areas?"

An efficient method to identify the critical points in a complex repair network involves the use of a Markov model of flows through the repair system. Initial work to apply this technique to depot repair processes was done by Peter S. Palmer (1970) of Charles Stark Draper Lab and by Russell M. Genet (Oct., 1970) of the Aerospace Guidance and Metrology Center. Application of this approach to the KT-73 IMU repair process is reported by Watson and Waterman (1974). Iwerson, Berry, and Brawner (1975) then extended the Watson-Waterman model to provide a more detailed description of the KT-73 repair system. The Markov modeling technique has also been applied by Genet, Martin, and Besteda (May, 1973) to depot repair processes for the G-200 Gyroscope and for the LN-12 inertial platform.

The objective of the above models is to describe the primary flow and cost characteristics of the repair process, including "feed back" characteristics. The value of making specific changes in the system may then be determined by resolving the model using changed parameters, and comparing the average repair costs under the old and proposed systems.

The basic mathematical structure of these models is as follows:

Let

Π_j = The average number of times that an IMU passes through repair stage j

C_j = Cost of passing through repair stage j

P_{ij} = Probability that an item progresses from repair stage i to repair stage j

The values for Π_j are the solution to the system of equations

$$\Pi_1 = 1 \qquad (90)$$

and

$$\Pi_j = \sum_i \Pi_i \cdot P_{ij} \qquad j = 2,3,\ldots,N \qquad (91)$$

where Π_1 represents the receiving stage for the process. Equation (90) simply states that one unit enters the repair process, while equation (91) states that the average number of times an item passes through stage j equals the sum, over all repair stages i, of the number of times an item passes through stage i times the probability the item progresses from stage i to stage j.

The average cost (C) to repair a single unit is then

$$C = \sum_j C_j \Pi_j \qquad (92)$$

For equations (90) and (91) to be valid, we must assume that the transition probabilities P_{ij} are dependent upon the current repair stage i, and do not depend upon the previous repair history of the item. This is known as the "Markov Property" assumption (Wagner, 1969, p. 740).

Let us now examine how this model may be used to identify critical areas in the KT-73 repair process.

Figure 19 illustrates the Watson-Waterman Model for the KT-73 repair process, with all feedback paths identified. In the figures, each repair stage is identified by a single letter and each flow path is represented by two letters. For example, the flow path AC represents flow from the initial test (stage A) to the Remove and Replace Gimbol Cluster Assembly stage (stage C). Transition probabilities for each flow stage are shown in Figure 20. Since there are N=20 stages, in Figure 19, equations (90) and (91) describing Figure 19 will consist of 20 equations. For example, the first three of these are

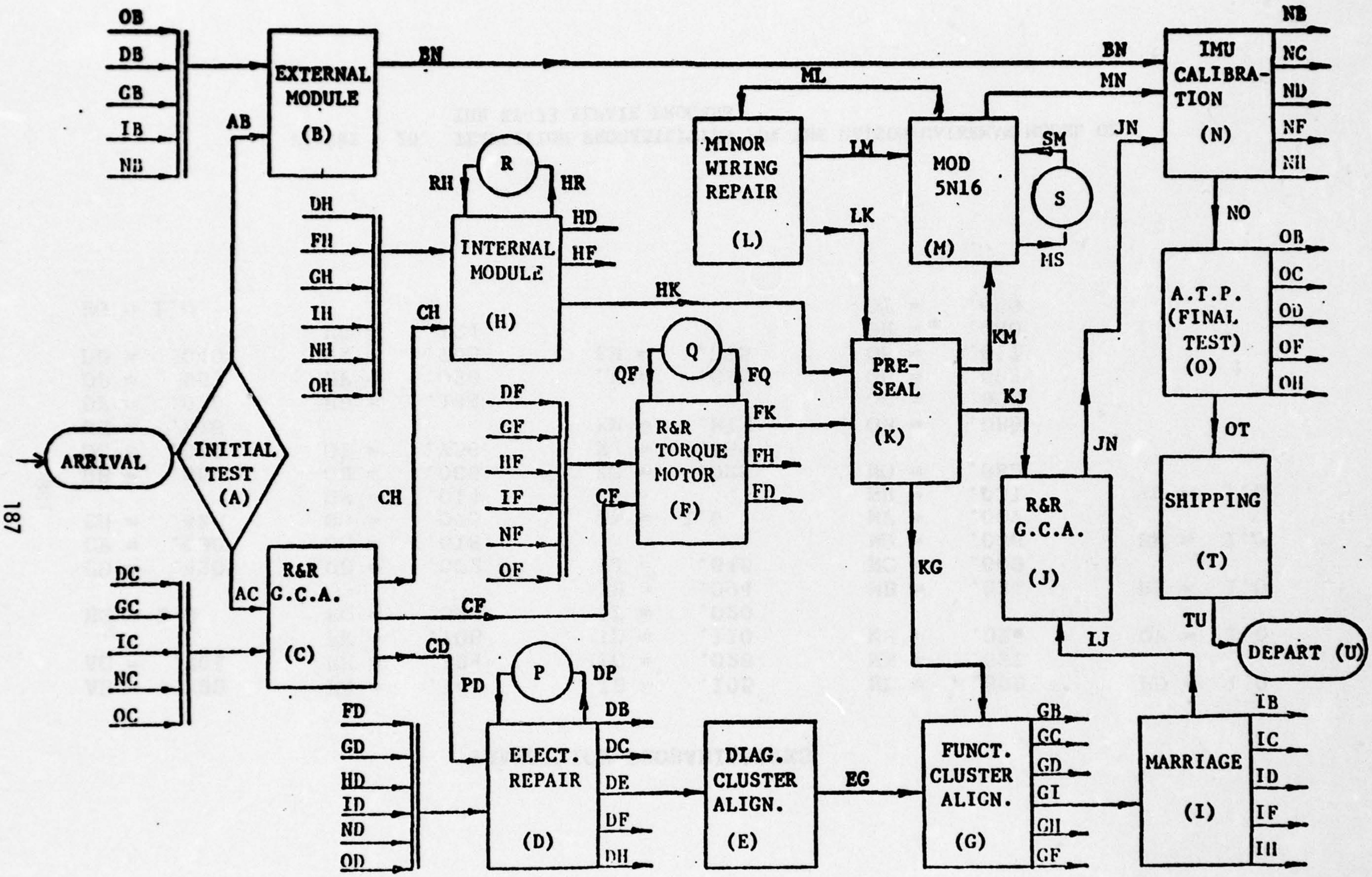


FIGURE 19. KT-73 REPAIR PROCESS SHOWING ALL FEEDBACK PATHS

TRANSITION PROBABILITIES

AB = .799
AC = .201

BC = 1.0

CD = .490
CF = .090
CH = .420

DB = .054
DC = .014
DE = .816
DF = .010
DP = .057
DH = .049

EG = 1.0

FD = .144
FH = .124
FK = .706
FQ = .026

GB = .072
GC = .018
GD = .075
GF = .014
GH = .065
GI = .756

HD = .144
HF = .026
HK = .706
HR = .124

IB = .105
IC = .026
ID = .110
IF = .020
IH = .094
IJ = .645

JN = 1.0

KG = .023
KJ = .159
KM = .818

LK = .615
LM = .385

ML = .309
MN = .667
MS = .024

NB = .034
NC = .009
ND = .036
NF = .007
NH = .031
NO = .883

OB = .089
OC = .022
OD = .093
OF = .017
OH = .080
OT = .699

PD = 1.0

QF = 1.0

RH = 1.0

SM = 1.0

TU = 1.0

FIGURE 20. TRANSITION PROBABILITIES FOR THE WATSON-WATERMAN MODEL OF THE KT-73 REPAIR PROCESS.

$$\Pi_A = 1$$

$$\Pi_B = \Pi_A P_{AB} + \Pi_D P_{DB} + \Pi_G P_{GB} + \Pi_I P_{IB} + \Pi_N P_{NB} + \Pi_O P_{OB}$$

$$\Pi_C = \Pi_A P_{AC} + \Pi_D P_{DC} + \Pi_G P_{GC} - \Pi_I P_{IC} + \Pi_N P_{NC} + \Pi_O P_{OC}$$

where P_{AB} denotes the probability that an item at stage A will be routed to stage B, and other P_{ij} values are similarly defined.

This model may be used to identify critical cost areas by performing the following steps.

a. First, solve the system of equations (90) and (91) for the values of Π_j. Since (90) and (91) is a system of N linear equations in N unknowns, the values of Π_j may be easily determined using readily available computer programs.

b. Given Π_j, the cost of stage j repair for an average (CS_j) item is then $CS_j = C_j \Pi_j$, and the cost to repair an average item is given by (92). Repair stages for which CS_j is large represent high impact stages; if the cost of such stages can be reduced, significant savings will be produced for the whole system.

c. Finally, investigate the impact of specific feedback components upon the average cost of repair. To do this, select a specific stage for further study. Suppose we select stage O, the A.T.P. Final Test stage. Next, reduce the feedback transition probabilities by a reasonable amount, say 10%, with a corresponding increase in the feed-forward path. Since we have selected stage O, we would reduce the transition probabilities for the feedback paths OB, OC, OD, OF, and OH, and correspondingly increase the transition probability for path OT. The system (90) and (91) would then be solved using the new coefficients, and the average repair cost (92) for the modified system would be determined. The difference between this cost and the cost found in step b represents the marginal savings that can be achieved by reducing feedback at stage O. A similar analysis would then be performed for each repair stage in the network.

For a detailed illustration of this analysis approach, see Watson and Waterman (1974) or Iwerson, Berry, and Brawner (1975).

The above analysis process provides two useful kinds of information. Step b identifies high cost repair activities. Small percentage savings in the cost of these activities should produce significant dollar savings to the system as a whole.

On the other hand, step c identifies high cost error flows. Feedback results from one of two types of errors; errors due to improper repair at some stage or errors due to incorrect test or diagnosis decisions. Step c analysis identifies areas in which small percentage reductions in feedback flow may produce significant cost savings. Unfortunately, the analysis does not determine the source or cause of these flows. Nevertheless, step c analysis is important in that it identifies areas in which more detailed analysis may be warranted. For these critical areas, the more detailed analysis procedures described elsewhere in this report should be used.

BIBLIOGRAPHY

Astrom, K.J., "Optimal Control of Markov Processes with Incomplete State Information", J. Math. Anal. Appl. v. 10, 1965, p. 174-205.

Betaque, N.E. and G. Anthony Gorry, "Automating Judgemental Decision Making for a Serious Medical Problem", Management Science, v. 17, n. 8, April, 1971, p. B-421-434.

Brown, F.D., and N.F. McAllister, "An Application of Inverse Probability to Fault Isolation", IRE Transactions on Military Electronics, July 1962.

Eckles, James E., "Optimum Maintenance with Incomplete Information", Operations Research, v. 16, 1968, p. 1058-1067.

Firstman, Sidney I. and Brian Gluss, "Optimum Search Routines for Automatic Fault Location", Operations Research, v. 8, p. 512-525, July-August, 1960.

Genet, Russell M., Avionics Cost Reduction Through Improved Tests, Aerospace Guidance and Metrology Center (AFLC), Newark Air Force Station, Newark, Ohio, 5pp.

Genet, Russell M., "An Introduction to Repair Engineering and Its Application at Special Repair Facilities", Nov. 1969, pp. 44.

Genet, Russell M., A Nonlinear Programming Determination of Depot Maintenance Errors, Aerospace Guidance and Metrology Center, Newark, Ohio, 1 October 1970, 36 pp.

Genet, Russell M., Richard R. Martin, and Roosevelt O. Besteda, An Assessment of Cost and Reliability Changes in the LN-12 Inertial Platform (Project Rivot Gjrot), Aerospace Guidance and Metrology Center, Newark Air Force Station, Newark, Ohio, 30 May 1973, 81 pp.

Hillier, F.S. and G.L. Lieberman, Operations Research, Holden-Day, 1974, pp. 534-573.

Howard, Ronald A., Dynamic Probabilistic Systems: Vo. II: Semi-Markov and Decision Processes, John Wiley and Sons, 1971.

Iwerson, Alfred, R.R. Berry, and J.E. Browner, A Cost Analysis of the KT-73 Inertial Measurement Unit Repair Process Using Graphical Evaluation Review Technique Simulation, unpublished master's thesis, Air Force Institute of Technology (AFLC/SL), Wright Patterson AFB, Ohio 45433, 145 pp.

Jordine and Sibson, Mathematical Taxonomy, Wiley, 1971.

Kemeny, John G. and Laurie J. Snell, Finite Markov Chains, D. Van Nostrand Company, Princeton, N.J., February 1963, p. 43-48.

Satia, J.K. and R.E. Lave, "Markovian Decision Processes with Probabilistic Observation of States", Management Science, v. 20, n. 1, September 1973, p. 1-13.

Smallwood, R.D. and E.S. Sondik, "Optimal Control of Partially Observable Markov Processes Over a Finite Horizon", Operations Research, v. 21, 1973, pp. 1071-1088.

Wagner, Harvey M. "Dynamic Programming in Markov Chains", Chapt. 18, Principles of Operations Research, Prentice Hall, 1969, p. 739-765.

Watson, E.B. and C.R. Waterman, A Cost Analysis of the KT-73 Inertial Measurement Unit Repair Process Using Absorbing Markov Chains, (thesis), Air Force Institute of Technology, Wright Patterson AFB, Ohio, 1974, 115 pp.

White Chelsa C., "Procedures for the Solution of a Finite-Horizon, Partially Observed, Semi-Markov Optimization Problem", Operations Research, V. 24, n. 2, 1976, p. 348-358.

SOME ASPECTS OF STATISTICAL CLASSIFICATION METHODS

by

Jack C. Lee
Department of Mathematics
Wright State University
Dayton, Ohio

1. INTRODUCTION

Classification may be considered as a problem of statistical decision functions. Suppose an individual is an observation from one of the k populations $\pi_1, ---, \pi_k$. The classification of an observation depends on the Vector measurement $\underset{\sim}{x}' = (x_1, ---, x_p)$ of that individual. An optimal classification procedures is the one which minimizes the expected cost or the probability of misclassification.

If q_i *is the prior probability of drawing an observation from* population π_i with probability density $p_i(\underset{\sim}{x})$ $(i=1, ---, k)$ and if the cost of misclassifying an observation from π_i as from π_j is $C(j/i)$, then the regions of classification, $R_1, ---, R_k$ that minimizes the expected cost are defined by assigning $\underset{\sim}{x}$ to R_m if

$$(1.1)\quad \sum_{\substack{i=1\\i\neq m}}^{k} q_i\, p_i\,(\underset{\sim}{x})\; C\,(m|i) < \sum_{\substack{i=1\\i\neq j}}^{k} q_i\, p_i\,(\underset{\sim}{x})\; C(j|i)\;(j=1, ---, k;\; j=m),$$

See Anderson (1958). We note in passing that when $k=2$, then we assign $\underset{\sim}{x}$ to R_1 if $q_2p_2(\underset{\sim}{x})\, C(1|2) < q_1p_1(\underset{\sim}{x})\, C(2|1)$.

Statistical Classification methods can also be applied to medical diagnosis, as in this situation we have essentially several possible populations for any patient to be classified, see, for example, Cornfield,

et al (1973) or Anderson (1973). With the help of computer, statistical classification methods should play more important role in the field of medical diagnosis for years to come.

In the next few sections we will discuss problems of dimensionality, multivariate normal classification, equal-mean classification and growth curve classification. We will restrict our attention to the situation where normality is assumed.

2. REDUCTION OF DIMENSIONALITY

It is sometimes of interest to determine whether we can reduce the dimensionality of observations from p variables to r $(<p)$ variables and preserve as much as possible the allocation in r-dimensions. This is, of course, the problem of principal component.

Let $\underset{\sim}{X}_1, ---, \underset{\sim}{X}_N$ be N independent and identically distributed p-variate normal with mean vector $\underset{\sim}{\mu}$ and Covariance Matrix Σ. Also let $\ell_1 >--- > \ell_p$ be the eigenvalues of $S=(s_{ab}) = \sum_{i=1}^{N} (\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})'$, $N\bar{\underset{\sim}{X}} = \sum_{i=1}^{N} \underset{\sim}{X}_i$, whereas $\lambda_1 >--->\lambda_p$ are the latent roots of Σ. One possible statistic to use for the reduction of dimensionality is $R = \sum_{i=1}^{r} \ell_i \Big/ \sum_{i=1}^{p} \ell_i$ which was proposed by Rao (1964). We will state a recent result of Krishnaiah and Lee (1976) which includes R as a special case.

Let

$$(2.1) \quad T_g = \sqrt{n}\,\{ \sum_{j=1}^{p} c_{gj}\, \ell_j \,(\sum_{h=1}^{p} d_{gh}\, \ell_n)^{-1} - \sum_{j=1}^{p} c_{gh}\, \lambda_j \,(\sum_{h=1}^{p} d_{gh}\, \lambda_n)^{-1} \}$$

for $g=1, 2, ---, q$, and c_{gh}, d_{gh} are known constants. We note that if $q=1$, $d_{gh}=1$ for all h,

$$C_{gh} = \begin{cases} 1 & \text{for } h = 1,2, ---, r \\ 0 & \text{for } h = r+1, ---, p \end{cases}$$

then we have essentially the statistic R proposed by Rao.

The joint density of $T_1, T_2, ---, T_q$ is

$$(2.2)\quad f(T_1, ---, T_q) = N(\underset{\sim}{T}; B_3^{-1})[1 + \frac{1}{\sqrt{n}}\{\sum_{h=1}^{p}\sum_{k\neq h}^{p}\sum_{j=1}^{q} A^*_{jh}\ \lambda^{-1}_{hk}\ \lambda_h\ \lambda_k\ H_j\ (\underset{\sim}{T})$$

$$-2\sum_{h=1}^{p}\sum_{j=1}^{q} b^*_{jh}\ \lambda_h^2\ H_j\ (\underset{\sim}{T})$$

$$+\frac{4}{3}\sum_{h=1}^{p}\sum_{j_1=1}^{q}\sum_{j_2=1}^{q}\sum_{j_3=1}^{q} \lambda_h^3\ A^*_{j_1h}\ A^*_{j_2h}\ A^*_{j_3h}\ H_{j_1j_2j_3}\ (\underset{\sim}{T})$$

$$-4\sum_{h=1}^{p}\sum_{h'=1}^{p}\sum_{j_1=1}^{q}\sum_{j_2=1}^{q}\sum_{j_3=1}^{q} \lambda_h^2\ \lambda_{h'}^2\ A^*_{j_1h}\ A^*_{j_2h}\ b^*_{hh',j_3}\ H_{j_1j_2j_3}\ (\underset{\sim}{T})$$

$$+ O(n^{-1})]$$

Where

$$B_3 = 2\sum_{h=1}^{p} \lambda_h^2\ A^*_h\ A^{*\prime}_h,\ A^*_h = (\Lambda_1^2\ A_{1h}, ---, \Lambda_q^2\ A_{qh}),\ \underset{\sim}{T}' = (T_1, ---, T_q)$$

$$\Lambda_g^{-1} = \sum_{h=1}^{p} d_{gh}\lambda_h,\ A^*_{gh} = \Lambda_g^2\ A_{gh},\ \lambda_{hk} = \lambda_h - \lambda_k,\ b^*_{gj} = \Lambda_j^3\ A_{gj}\ d_{gj},$$

$$b^*_{gf,j} = \Lambda_j^3\ A_{jg}\ d_{jf},\ N(x;\Omega) = (2\pi)^{-p/2}\ |\Omega|^{-\frac{1}{2}}\ \exp[-\tfrac{1}{2}\ \underset{\sim}{x}'\ \Omega^{-1}\ \underset{\sim}{x}],$$

i.e. the density of $N(0,\Omega)$, and $H_{j_1,---,j_s}(\underset{\sim}{x}) = \dfrac{(-1)^s\ \partial^s}{N(\underset{\sim}{x};\Omega)\ \partial x_j\ ---\partial x_{j_s}}\ N(\underset{\sim}{x};\Omega)$

where $j_1, j_2, ---, j_s$ are s integers such that $1 \leq j_i \leq p$.

$H_{j,---,j_s}(\underset{\sim}{x})$ is the multivariate Hermite Polynomials, see Appel and Kampe de Feriet (1926), and Khatri and Mitra (1969).

With (2.2) we will be able to test various hypotheses and obtain confidence intervals (regions) for functions of eigenvalues $\lambda_1, ---, \lambda_p$. In particular, the hypothesis testing as well as confidence interval for

$$\sum_{i=1}^{r} \lambda_i \quad \sum_{j=1}^{p} \lambda_j$$

can be obtained. Results for the joint distribution of linear combinations of eigenvalues and those of the corresponding complex multivariate normal case are also given by Krishnaiah and Lee (1976).

In order to make practical use of (2.2) and the other results derived by Krishnaiah and Lee (1976), estimates would have to be substituted for the λ_i, which was also the situation reported by Lawley (1956) in considering tests for the equality of last few latent roots. Certainly better results would be the ones free from any unknown parameters. Unfortunately they are not yet available.

3. MULTIVARIATE NORMAL CLASSIFICATION

Suppose there are k p-variate normal populations $\pi_1, ---, \pi_k$ with mean vectors $\underset{\sim}{\mu}_1, ---, \underset{\sim}{\mu}_k$ and covariance matrices $\sum_1, ---, \sum_k$. One is interested in classifying a new p-variate observation $\underset{\sim}{z}$ to one of these possible populations in some optimal fashion. Assuming $\underset{\sim}{z}$ has prior probability q_i of belonging to π_i, $\sum_{i=1}^{k} q_i = 1$, then the optimal classification method with regard to posterior probability of correct classification is to allocate $\underset{\sim}{z}$ to that π_i for which

$$(3.1) \quad A_i(p) = \ln q_i - \tfrac{1}{2}\ln |\textstyle\sum_i| - \tfrac{1}{2}(\underset{\sim}{z}-\underset{\sim}{\mu}_i)' \textstyle\sum_i^{-1} (\underset{\sim}{z}-\underset{\sim}{\mu}_i)$$

is a maximum. Here, of course, the parameters $\sum_i$, $\underset{\sim}{\mu}_i$ and q_i are assumed known.

Geisser (1964, 1966) proposed the following Bayesian method. Suppose π_i has density $f_1(\cdot|\theta_i,\psi_i)$ where θ_i is the set of unknown parameters and ψ_i is the set of known parameters; X_i are the data obtained from π_i based on N_i independent vector observations. Further, let $\theta = \bigcup_{i=1}^{k} \theta_i$, $\psi = \bigcup_{i=1}^{k} \psi_i$, $g(\theta|\psi)$ be the joint prior density of θ, and $L(X_i|\theta_i, \psi_i)$ be the likelihood fo the sample obtained from π_i with joint likelihood.

$$(3.2)\quad L(X|\theta,\psi) = \prod_{i=1}^{k} L(X_i|\theta_i,\psi_i)$$

where X represents the totality of samples $X_1, ---, X_k$. The posterior probability that $\underset{\sim}{z}$ belongs to π_i is

$$(3.3)\quad P_r(\underset{\sim}{z}\varepsilon\pi_i|X,\psi,\underset{\sim}{q}) \propto q_i f_2(\underset{\sim}{z}|X,\psi,\pi_i)$$

where $q' = (q_1, ---, q_k)$ and

$$(3.4)\quad f_2(\underset{\sim}{z}|X,\psi,\pi_i) = \int f_1(\underset{\sim}{z}|\theta_i,\psi_i,\pi_i)\, P(\theta_i|X,\psi)\, d\theta_i,$$

$$(3.5)\quad P(\theta_i|X,\psi) = \int P_1(\theta|X,\psi)\, d\theta_i^c \propto \int L(X|\theta,\psi)\, g(\theta|\psi)\, d\theta_i^c$$

where θ_i^c is the complement of θ_i, $\theta_i^c \cup \theta_i = \theta$. It is to be noted that $f_2(\underset{\sim}{z}|X, \psi, \pi_i)$ is the predictive density of $\underset{\sim}{z}$ and $P(\theta_i|X,\psi)$ is the posterior density of θ_i. For classification purposes, we may choose to assign $\underset{\sim}{z}$ to that π_i for which (3.3) is a maximum. We thus see that all we need to derive is the predictive density of $\underset{\sim}{z}$. When normality and parameters are assumed, then (3.3) and (3.1) are equivalent. We also note that in (3.5) relevant constant of proportionality has to be recovered, otherwise it might be too complicated to recover in the final result, see Lee (1975a).

Geisser (1964), using noninformative prior for unknown parameters derived predictive densities for various combinations of mean vectors and covariance matrices. For the rest of this paper we will discuss equal-mean classification and that of growth curves.

4. EQUAL-MEAN CLASSIFICATION

The problem of statistical discrimination with respect to different unknown covariance matrices and common unknown mean vector of two p-variate normal populations π_1, π_2 was studied by Okamoto (1961), Geisser (1964) and Lee (1975a).

Let $\underset{\sim}{\mu}$ be the common unknown mean vector and $\sum_1, \sum_2$ the two unknown covariance matrices. By using a convenient joint prior

$$g(\underset{\sim}{\mu}, \textstyle\sum_1^{-1}, \sum_2^{-1}) \propto |\textstyle\sum_1^{-1}|^{(p+1)/2} \, |\textstyle\sum_2^{-1}|^{(p+1)/2}$$

Lee (1975a) derived the predictive density

$$(4.1)\quad f_3(\underset{\sim}{z}|X, \pi_i) \propto \frac{1}{\Gamma[(N_j+1-p)/2]\,\Gamma[N_j/2]} |A_{iz}|^{-N_i/2} |A_j|^{-(N_j-1)/2}$$

$$\cdot \int_0^\infty t^{(N_i-1)/2} B_z^{-(N+1-p)/2} |N_j A_{iz} + t(N_i+1)A_j|^{-\frac{1}{2}} dt,$$

$i \neq j$, where $N_i \underset{\sim}{\bar{X}}_i = \sum_{d=1}^{N_i} \underset{\sim}{X}_{id}$, $A_i = \sum_{d=1}^{N_i} (\underset{\sim}{X}_{id} - \underset{\sim}{\bar{X}}_i)(\underset{\sim}{X}_{id} - \underset{\sim}{\bar{X}}_i)'$,

$A_{iz} = A_i + \frac{N_i}{N_i+1} (\underset{\sim}{z}-\underset{\sim}{\bar{X}}_i)(\underset{\sim}{z}-\underset{\sim}{\bar{X}}_i)'$, $\underset{\sim}{\bar{\mu}} = (N_i+1)^{-1} (N_i \underset{\sim}{\bar{X}}_i + \underset{\sim}{z})$,

$B_z = 1 + t[1+N_j(N_i+1)(\underset{\sim}{\bar{\mu}}-\underset{\sim}{\bar{X}}_j)'\{N_j A_{iz} + t(N_i+1)A_j\}(\underset{\sim}{\bar{\mu}}-\underset{\sim}{\bar{X}}_j)]$

and $X_{i1}, ---, X_{iN_i}$ are N_i independent observations from π_i. It is to be noted that the proportionality constant is irrelevant for our purpose.

A computer program written in Fortran was developed for the evaluation of the above integral. We also note that in the integral we have for any t^*

$$(4.2)\quad \int_{t^*}^{\infty} h(t)\,dt < 2N_j^{-p/2}\ (N_j-p)^{-1}\ t^{*-(N_j-p)/2}\ |A_{iz}|^{-\frac{1}{2}} = E_1$$

where $h(t)$ is the integrand of (4.1). Thus the integral in (4.1) can be considered as a finite integral from 0 to t^* with error bound E_1. The computer program also provides error bound E_2 of integration from 0 to t^*. Hence the error bound of the integral in (4.1) is simply $E_1 + E_2$.

5. GROWTH CURVE CLASSIFICATION

The growth curve model introduced by Potthoff and Roy (1964) is

$$(5.1)\quad E(X_{p\times m}) = B_{p\times m}\,\tau_{m\times r}\,A_{r\times N}$$

where τ is unknown, B and A are known design matrices of ranks $m < p$ and $r < N$ respectively. Further, the columns of X are independent p-dimensional normal variates having a common unknown covariance matrix $\sum$, i.e. $G(X|\tau,\sum) = N(\cdot;B\tau A, \sum \otimes I_N)$ where $\otimes$ denotes the Kronecker product and $G(\cdot)$ is the c.d.f.

We will consider the situation where we have observed k growth curves

$$(5.2)\quad G(X_i|\tau_i,\textstyle\sum_i,\pi_i) = N(\cdot;\ B\tau_i A_i,\ \textstyle\sum_i \otimes I_{Ni}),\ i=1, ---, K$$

and a future observation $\underset{\sim}{z}$, of dimension $p \times 1$, which is known to be drawn from one of the K populations π_1, ---, π_k with known prior probabilities q_1, --- q_k. We assume $G(\underset{\sim}{z}|\tau_i,\sum_i, \pi_i) = N(\cdot;\ B\tau_i F_i,\ \sum_i)$ where F_i is a known vector.

By a convenient join prior

$$(5.3) \quad G(\tau_1, ---\tau_k, \textstyle\sum_1^{-1}, ---, \sum_k^{-1}) \propto \prod_{\alpha=1}^{K} |\textstyle\sum_\alpha|^{(p+1)/2}$$

the predictive density of $\underset{\sim}{z}$ was shown by Lee (1975b) to be

$$(5.4) \quad f_4(\underset{\sim}{z}|X_i, \pi_i) \propto C_i^* \, |B' S_i^{-1} B|^{-(N_i-r)/2} \, |G_i|^{m/2} \, |Z' X_i X_i' Z|^{(N_i-m)/2}$$

$$\cdot |(B' S_i^{-1} B)^{-1} + B^*(\underset{\sim}{z}-BQ_i) \, G_i (\underset{\sim}{z}-BQ_i)' \, B^{*\prime}|^{-(N_i+1-r)/2}$$

$$\cdot |Z'(X_i X_i' = \underset{\sim\sim}{zz}') \, Z|^{-(N_i+1-m)/2}$$

$$\text{for } i = 1, ---, k$$

where

$$G_i^{-1} = M_i^{-1} + (\underset{\sim}{z}-\hat{V}_i)' \, Z(Z' S_i Z)^{-1} \, Z'(\underset{\sim}{z}-\hat{V}_i)$$

$$M_i = 1-F_i'(H_i H_i')^{-1} \, F_i, \; H_i = (A_i, F_i)$$

$$\hat{V}_i = X_i A_i' \, (A_i A_i')^{-1} \, F_i, \; B^* = (B'B)^{-1} B'$$

$$(5.5) \quad Q_i = B^* \hat{V}_i + B^* S_i Z(Z' S_i Z)^{-1} \, Z'(\underset{\sim}{z}-\hat{V}_i)$$

$$S_i = X_i (I-A_i'(A_i A_i')^{-1} \, A_i) \, X_i'$$

$$C_i^* = \frac{\Gamma[\frac{1}{2}(N_i+1-r)] \; \Gamma[\frac{1}{2}(N_i+1-m)]}{\Gamma[\frac{1}{2}(N_i+1-r-m)] \; \Gamma[\frac{1}{2}(N_i+1-p)]}$$

and Z is $p \times p\text{-}m$ and of rank $p\text{-}m$ such that $B'Z = 0$.

The constant of proportionality is irrelevant and will be absorbed by the proportionality sign $\propto$.

The predictive densities for various combinations of parameters together with a special covariance structure were given by Lee (1975b). In case locational structure is present, the growth curve classification is expected to yield result better than the general multivariate normal classification.

REFERENCES

1. Anderson, T. W. (1958). An Introduction to Multivariate Statistical Analysis. Wiley, New York.

2. Anderson, J. A. (1973). Logistic Discrimination with Medical Applications. In Discriminant Analysis and Applications, T. Cacoullos, ed. Academic Press, New York.

3. Appel, P. and J. Kampe de Feriet (1926). Functions Hypergeometriques et Hypersheriques. Gauthier-Villars, Paris.

4. Cornfield, J., R. A. Dunn, C. D. Batchlor and H. V. Pipberger (1973). Multigroup Diagnosis of Electrocardiograms. Computers and Biomedical Research, 6:97-120.

5. Geisser, S. (1964). Posterior Odds for Multivariate Normal Classifications. J.R.S.S. B, 26:69-76.

6. ____________(1966). Predictive Discrimination. In Multivariate Analysis, P. R. Krishnaiah, ed., 149-63, Academic Press, New York.

7. Khatri, C. G. and S. K. Mitra (1969). Some Identies and Approximations Concerning Positive and Negative Multinomial Distributions. In Multivariate Analysis - II, P. R. Krishnaiah, Ed., 241-60, Academic Press, New York

8. Krishnaiah, P. R. and J. C. Lee, (1976), Inference on the Eigenvalues of the Covariance Matrices of the Real and Complex Multivariate Normal Populations. In Multivariate Analysis IV, P. R. Krishnaiah, ed., North-Holland, New York (In press)

9. Lawley, D. N. (1956). Tests of Significance for the Latent Roots of Covariance and Correlation Matrices. Biometrika, 43:128-36.

10. Lee, J. C. (1975a). A Note on Equal-Mean Discrimination. Communications in Statistics, 4:251-4.

11. __________(1975b). Classification and Estimation of Growth Curves, A Bayesian Approach. ARL TR 75-0025.

12. Okamoto, M. (1961). Discrimination for Variance Matrices. Osaka Math. J., 13:1-39.

13. Potthoff, R. R. and S. N. Roy, (1964). A Generalized Multivariate Analysis of Variance Model Useful Especially for Growth Curve Problems. Biometrika, 51:313-26

14. Rao, C. R. (1964). The Use and Interpretation of Principal Component Analysis in Applied Research. Sankhyā A, 26:329-58.

DIAGNOSIS OF RELIABILITY REPAIR STAGE AND REMEDIES

Edward Bilikam and Albert H. Moore

Abstract

There are three stages of reliability failure patterns, (1) infantile failure, (2) cyclic failure and (3) fatigue failure. These stages determine the general method and condition of the replaceable units. Some remedies are burn-in, quality control, parts control and scheduled maintenance (overhauls). These measures can be applied according to the stage of the reliability failure pattern. The failure patterns should be determined by reliability testing and subsequent analysis. In the analysis of the data it is necessary to determine what stage you are in. One way of doing this is to use a Likelihood-Ratio Test. Such a test is presented with discussion of the power of the test. This test also has applications to reliability growth.

INTRODUCTION

We will be interested in the failure of equipment which will in general be replaced with repaired equipment. The fact that the equipment is old enough to have been repaired means that we are not just considering initial performance reliability but reliability of maintained equipment.

A way of characterizing the reliability of maintained equipment is the hazard rate curve. Hazard rate curves can be classified into three general patterns. The first pattern is that of infantile failure. In this stage the failure rate is much higher than after the failure rate has stabilized and is characterized by a sharply decreasing failure rate. Another pattern is characterized by cyclic failure. The installed equipment in this phase has an operating time before failure about equal to the MTBF and varies randomly on either side. The cyclic time of replacement is very nearly the MTBF. The stage is also characterized by a slowly increasing failure rate. The number of replacements in a fixed unit of time is the time divided by the MTBF. The third phase of the failure pattern is the fatigue phase. Here the equipment operates until the fatigue hazard sets in and the equipment exhibits a sharply increasing failure rate.

Remedies - Infantile Failure

The above description of infantile failure leads to a diagnosis of a producibility problem. Something must be

done to enable the operation of the equipment for a useful period (see reference 1). One remedy is burn-in. Just how long the burn-in must be applied to each equipment was discussed in reference [2]. Burn-in however can be expensive and/or destructive. For instance if it takes one hundred equipments to obtain one equipment that will survive to the operation stage then burn-in is very expensive indeed. It is even more critical if the burn-in leads to destructive failures. Then the equipment lost during burn-in just increases the cost per item. This is why burn-in of microcomponents is usually feasible while burn-in of sub-systems is usually prohibitive.

On the component level two other remedies may be useful. These are parts control and quality control. The natural laws of failure determine which is more useful. Quality control is applied to production equipment parts when it is determined that the quality of the parts can be maintained within engineering tolerances with only an occasional chance defect. Parts control however means screening the parts. That is weeding out the bad parts in selective inspection to obtain preferred classes of parts to use in critical assemblies. The idea is hold equipment performance at a minimal but necessary level.

When none of the above remedies can alone improve infantile failure there is a definite need for reliability growth. What is needed is a reliability growth into the

class of cyclic failure patterns. Redesign would then be a real consideration.

Remedies - Cyclic Failure

In cyclic failure the producibility of the equipment is usually not in question. The problem, if there is a problem is low MTBF. In general this calls for an increase in the scale parameter of the failure time distribution. Failures are occurring just too soon after the equipment is put into operation. The solution is to lengthen the equipment life. Redundancy and fault tolerance are considerations. That is one should try to rearrange and work with the base equipment that is on hand. Software redundancy for processors should be considered.

If it can be shown that cyclic failure is present, the next life step is fatigue. Actually fatigue failure is unwanted. The kind of repair for cyclic failure is unscheduled and preventive maintenance. That is "beef-up" the system. This may be difficult in airborne systems where weight and cost are critical. Engineering design is sometimes load critical so that when preventive maintenance is applied, the addition of new components changes the loading in an inappropriate way. The results of "improvements" must be tested (reliability environmental test) to see the result of changing the system.

Remedies of Fatigue

Some components are always subject to fatigue. Some relays are an example. There is little choice of a remedy

for fatigue-retrofit and scheduled maintenance. When components are worn out they must be replaced. When the whole system is in fatigue, scheduled overalls are a must. Fatigue is characterized by a gradual degeneration until a critical fatigue life where catastropic failure sets in.

Identification of Life Stages

The stages of failure patterns discussed above are characterized by the shape of the failure probability density function (p.d.f.). This function formulates quantitatively where the failures are most probable, in early failure, centralized failure or wear out failures. The class of Weibull distributions with different shape parameters will be used to model the failure distributions. The Weibull p.d.f. has the form

$$f(x) = (B/\theta)x^{B-1} \exp -(x^B/\theta) \qquad x, \theta, B > 0 \qquad (1)$$

An additional parameter to locate the distribution is obtained by the transformation $x = t - c$, where no failures occur with probability one before $c \geq 0$. Generally we will assume c is known or zero, but c may be estimated (see Harter and Moore). B is the shape parameter and θ is the B^{th} power of the scale parameter. Given the Weibull density (c=0) the reliability function (probability of survival) is

$$R(t) = \exp - (t^B/\theta) \qquad (2)$$

for operation time t.

Each of the parameters may be estimated from suitable sample data. Two common types of censored data treated in the literature are called Type I and Type II censoring. Type II censoring occurs when n equipments are put on test and the test is continued until r items have failed ($r \leq n$). In Type I censoring the test is terminated at a fixed time regardless of the number of items that have failed. Harter and Moore [4] using the method of maximum-likelihood have derived simultaneous nonlinear equations for the parameters of two and three parameter Weibull distribution and iterative techniques for their solution. Wingo [7] presented a way of accelerating the solution of the simultaneous equations using a gradient technique. For multiple independent Type I censored samples with failure times known and unknown Bilikam and Moore have used maximum likelihood techniques and the gradient iterative method to solve for θ and B. Therefore estimation techniques exist to estimate the parameters of the Weibull distribution for type of data considered above.

The stages of the failure patterns discussed above may be more precisely classified by the shape parameter of the Weibull distribution as follows:

a. $0 < B < 1.0$ infantile failure (see Figure 1)

b. $1.0 < B < 3.6$ gradual progression toward cyclic failure

c. $3.6 < B < 4.0$ cyclic failure (see Figure 2)

d. $B > 5$ fatigue failure (see Figure 3)

An estimation of the failure pattern within the scope of parametric statistics is possible by estimating the shape

Figure 1- INFANTILE FAILURE DISTRIBUTION

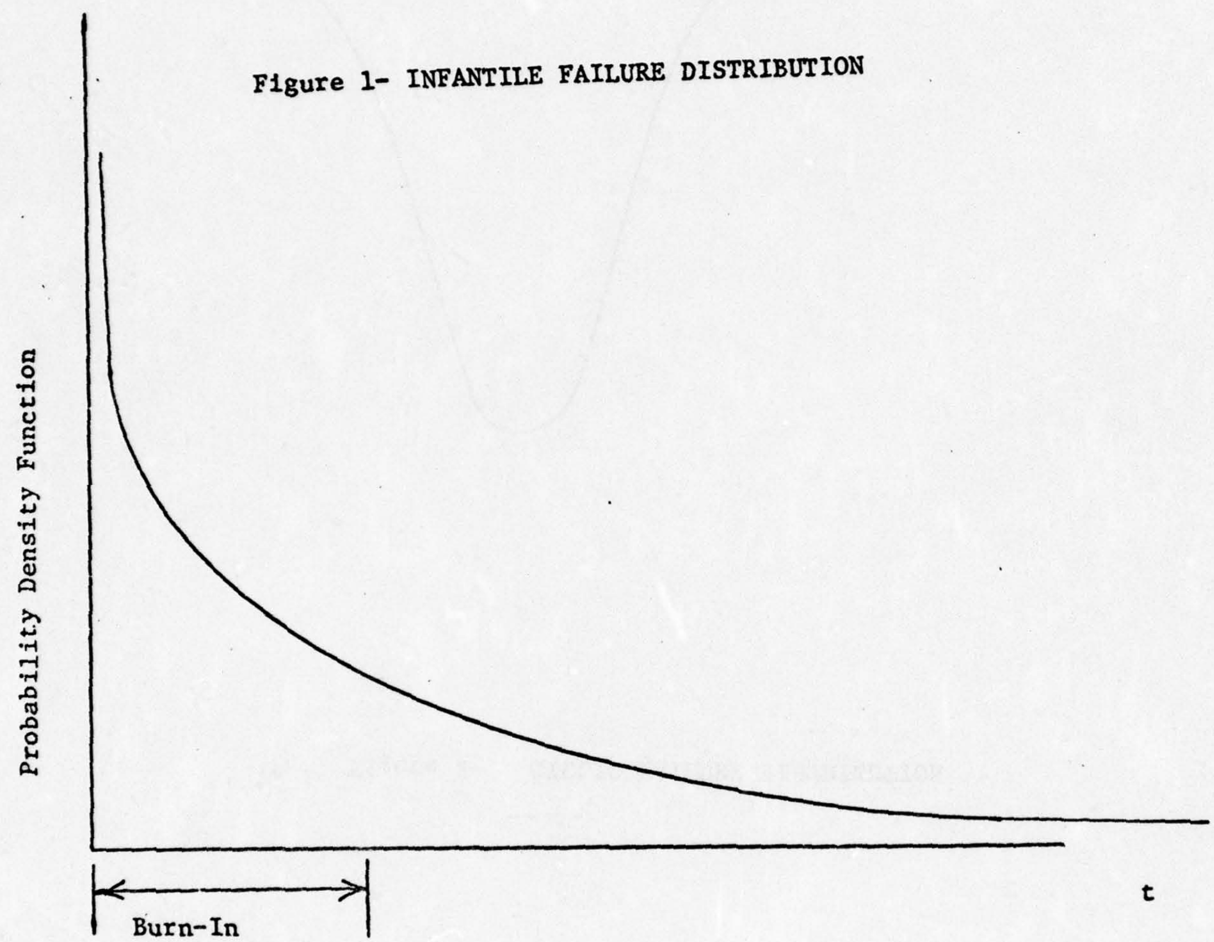
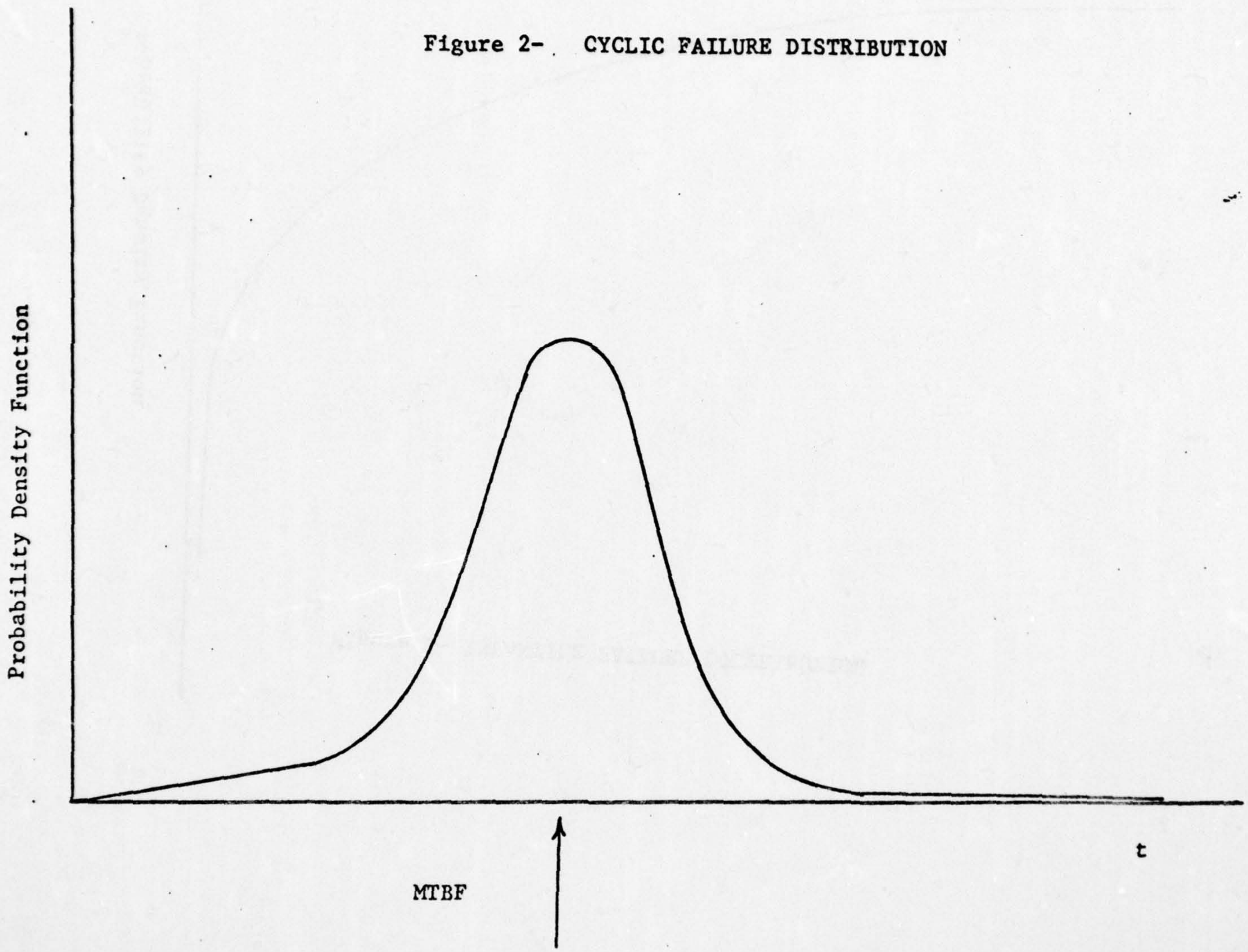


Figure 2- CYCLIC FAILURE DISTRIBUTION



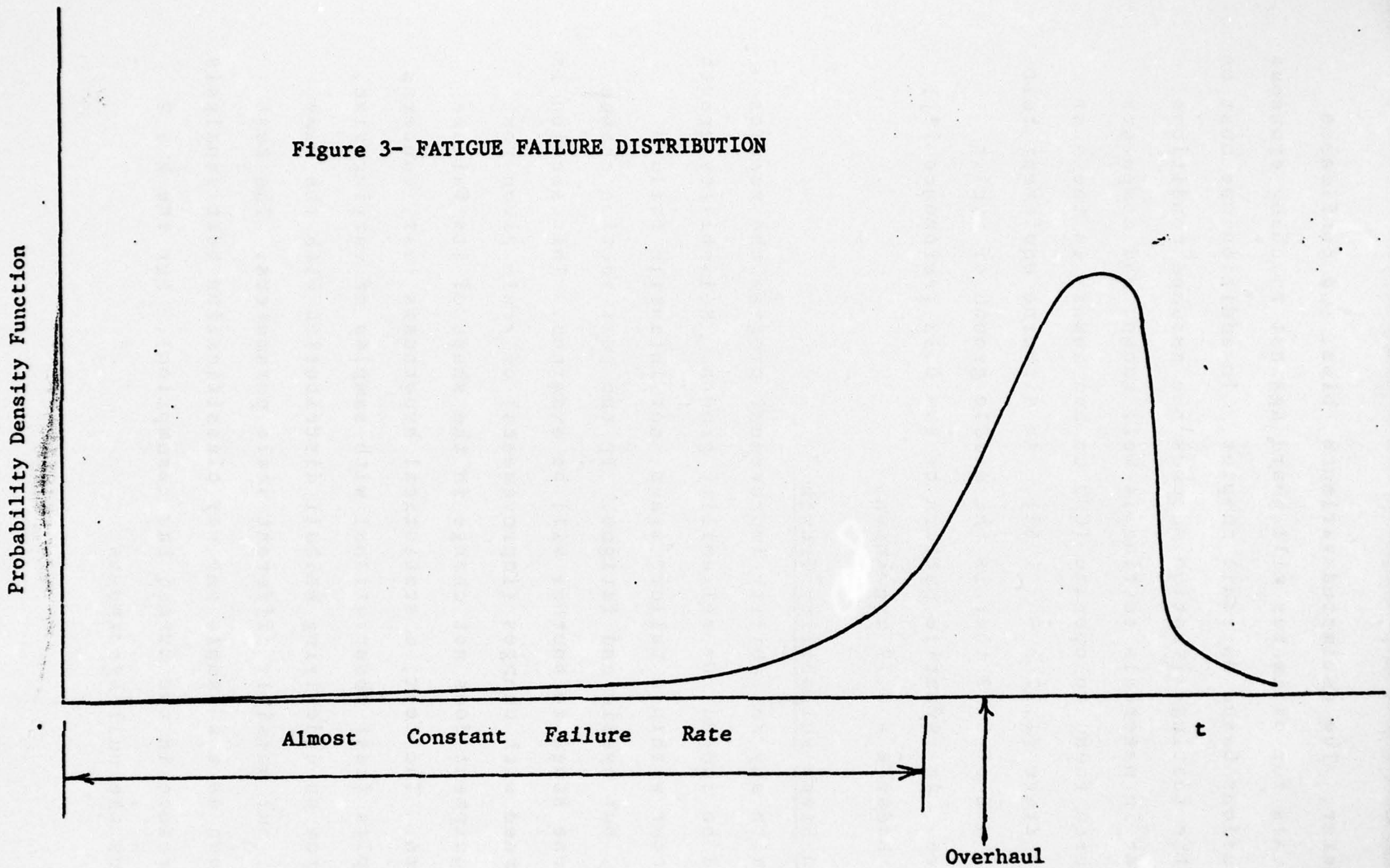


Figure 3- FATIGUE FAILURE DISTRIBUTION

parameter. The estimated variance, bias, and confidence intervals for estimates will guard against reaching eroneous conclusions based on small samples. In addition one must be sure the testing situation matches the assumed conditions. Fatigue in materials testing is well known, and computers have been found to operate (CPU up and downs) in the near cyclic stage ($1.12 < \hat{B} < 1.62$). An airborne equipment test showed a $\hat{B}$ of 2.22 that is the middle ground of cyclic failure. An infantile pattern of $\hat{B} = 0.32$ (reference [3]) and a Radar $\hat{B} = 0.72$ are known.

Within Stage Reliability Growth

With any reliability improvement program the results should be judged for reliability growth. Reliability growth can occur within a failure stage, not infantile failure stage, but cyclic and fatigue. In the next section change from one stage to another will be examined. This section is concerned with changes (improvements) of scale given that the equipment does not change in the shape of its failure pattern. The test, a statistical hypothesis test, concerns k samples (test observations) with samples of varying size, all from an underlying Weibull distribution with the same shape, but possibly different scale parameters. The test is known as a k-sample one way classification test (analysis of variance in the Normal Law assumption). For the $k \geq 2$ samples the null hypothesis

$$H_0: \eta = \eta_i \text{ (the k scale parameters}$$

$$\theta_i^{1/B}) \text{ for } i = 1,2,\ldots,k. \qquad (3)$$

That is, it is tested that, throughout the k reliability improvement tests there was no improvement in the scale. McCool [5] is a study of the statistical test applications with tabled results in which Type II censored samples are allowed. Of course, for reliability growth, there should be a significant deviation from hypothesis (3). Now the MTBF for a Weibull population is

$$\text{MTBF}_i = \eta_i \Gamma(1/B + 1)$$
$$i = 1,2,\ldots,k \qquad (4)$$

where

$$\eta_i = \theta_i^{1/B}. \qquad (5)$$

Hence testing that the η's are all equal from sample to sample, given the shape B does not change, tests that the MTBF's are not improving. The statistic used is a F ratio of the within group shape estimates to the pooled shape estimate.

Change of Failure Pattern

As was discussed above, the consequences of infantile failure are so severe that it is wished to know when this stage has been changed to cyclic failure, or normal maintenance conditions. Then too, for purposes of planning over-

hauls and retrofits, a change from cyclic failure to fatigue failure should be detected. These changes of stage must be observed from reliability test results. The two sample change of shape parameter is derived from the likelihood equations.

The null hypothesis is that there is no change of shape parameter B from the first sample (size n_1) to that of the second sample (size n_2) after possible improvement or change of shape due to overall effect of age.

$$H_0 : B_1 = B_2 \tag{6a}$$

The alternative hypothesis is unequal shape parameters

$$H_1 : B_1 \neq B_2 \tag{6b}$$

The test is independent of the scale parameter. The *maximum likelihood ratio statistic* is to be calculated from observed failure times of the first sample $x_1, \ldots, x_{n_1}$ and those of the second sample $y_1, \ldots, y_{n_2}$. $\hat{\lambda}$ was derived for the test by J. E. Bilikam.

$$\hat{\lambda} = (\hat{\hat{B}}_1/\hat{B}_1)^{n_1} (\hat{\hat{B}}_1/\hat{B}_2)^{n_2} \prod_{i=1}^{n_1} x_i^{\hat{\hat{B}}_1 - \hat{B}_1}$$

$$\prod_{j=1}^{n_2} \cdot \, y_j^{\hat{\hat{B}}_1 - \hat{B}_2} \, (\hat{\theta}_1/\hat{\hat{\theta}}_1)^{n_1} (\hat{\theta}_2/\hat{\hat{\theta}}_2)^{n_2} \tag{7}$$

where $\hat{\hat{B}}_1$, $\hat{\hat{\theta}}_1$, $\hat{\hat{\theta}}_2$ are M.L. estimates under H_0 and $\hat{B}_1$, $\hat{B}_2$, $\hat{\theta}_1$, $\hat{\theta}_2$ are general M.L. estimates for each sample. Under H_0:

$$\frac{n_1+n_2}{\hat{\hat{B}}_1} + \sum_{i=1}^{n_1} \ln x_i - x_i^{\hat{\hat{B}}_1} \frac{\ln x_i}{\hat{\hat{\theta}}_1} + \sum_{j=1}^{n_2} \ln y_j - y_j^{\hat{\hat{B}}_1} \frac{\ln y_j}{\hat{\hat{\theta}}_2} = 0 \quad (8)$$

and

$$\hat{\hat{\theta}}_1 = \sum_{i=1}^{n_1} x_i^{\hat{\hat{B}}_1}/n_1, \quad \hat{\hat{\theta}}_2 = \sum_{j=1}^{n_2} y_j^{\hat{\hat{B}}_1}/n_2$$

Under H_1:

$$n_1/\hat{B}_1 + \sum_{i=1}^{n_1} \ln x_i - \sum_{i=1}^{n_1} x_i^{\hat{B}_1} \frac{\ln x_i}{\hat{\theta}_1} = 0$$

and

$$\hat{\theta}_1 = \sum_{i=1}^{n_1} x_i^{\hat{B}_1}/n_1 \quad (9)$$

$$n_2/\hat{B}_2 + \sum_{j=1}^{n_2} \ln y_j - \sum_{j=1}^{n_2} y_j^{\hat{B}_2} \frac{\ln y_j}{\hat{\theta}_2} = 0$$

and

$$\hat{\theta}_2 = \sum_{j=1}^{n_2} y_j^{\hat{B}_2}/n_2. \quad (10)$$

Each set of the above equations (8), (9), and (10) are solved simultaneously to obtain MLE's. Equation (7) $\hat{\lambda} = L(\hat{\hat{\omega}})/L(\hat{\Omega})$ is the value of the maximum likelihood ratio, the distribution of which may be used to construct the power of this test. Under H_0, the critical points of the test are obtained by

Petrick [6]. It was also shown by Petrick that the critical values $\lambda_c(\alpha, n_1, n_2)$ are not a function of B_1, θ_1, θ_2 but only of the sample sizes n_1, n_2 and the significance level α. Some values of the critical value follow in Table I.

Table I

Critical Points of the Two Sample Test of Weibull Change in Shape Parameter

H_0: $B_1 = B_2$ vs. $B_1 \neq B_2$

(B_1, B_2, θ_1, θ_2 unknown)

(Petrick [6])

Sample Sizes		Significance	Critical Point
n_1	n_2	α	λ_c
10	10	0.05	0.1108
10	20	0.05	0.1154
20	20	0.05	0.1277
30	30	0.05	0.1303
10	20	0.01	0.0230
20	20	0.01	0.0264
30	30	0.01	0.0296

When $\hat{\lambda} < \lambda_c$ the hypothesis H_0: $B_1 = B_2$ is rejected. The power of the test, the probability of not making a type II error is then the probability of rejecting H_0: $B_1 = B_2$ when, indeed, H_1 is true; that is when there is a shape change.

The power was obtained by Monte Carlo simulation. The simulation required solution of (9) and (10), for equation (7) for each trial assuming values of B_1 and B_2.

To generate the random Weibull numbers the probability integral theorem implies a random number r on (0,1) is related to the Weibull number x,

$$x = -\eta[\ln(1-r)]^{1/B} \quad (11)$$

Substituting x_i and y_j in (7) with η_1 and η_2 respectively from (11) it is clear that $\hat{\lambda}$ does not depend on θ_1 and θ_2 (i.e., on η_1 and η_2). Hence the distribution of $\hat{\lambda}$ is independent of the scale parameters. Hence the power is independent of the scale parameters. Actually Petrick shows this only under $H_0: B_1 = B_2$ but it is easily seen to follow under $H_1: B_1 \neq B_2$.

The power of the test was analyzed from Monte Carlo simulation results and is tabled on the following page.

Summary and Conclusions

From the failure times observed during reliability tests and the scale and shape parameter estimates one can determine the life stage a repairable equipment is in.

Several remedies for infantile failure and fatigue failure are possible but not always applicable. Infantile failure usually indicates a producibility problem that usually calls for engineering changes and/or design fixes. Cyclic failure requires unscheduled maintenance. Fatigue failure requires overhauls and scheduled maintenance.

Table II

Power of the Two Sample Test of Weibull

Change in Shape Parameter

$$H_0: B_1 = B_2 \text{ vs. } H_1: B_1 \neq B_2$$

$$\alpha = 0.05$$

Sample Sizes n_1	n_2	Shape Parameters B_1	B_2	Simulation Trials	Power P(reject H_0)
10	20	0.5	1.0	1000	0.531*
10	20	0.5	2.5	1000	0.994
10	20	1.0	2.5	1000	0.768
10	20	2.0	4.0	1000	0.531*
10	20	3.5	6.0	1000	0.330
20	20	0.5	1.0	500	0.802*
20	20	1.0	2.0	500	0.802
20	20	2.0	4.0	500	0.802
20	20	4.0	8.0	500	0.802
20	20	8.0	16.0	500	0.802
30	30	0.5	1.0	500	0.942*
30	30	1.0	2.0	500	0.942
30	30	2.0	4.0	500	0.942
30	30	4.0	8.0	500	0.942
30	30	8.0	16.0	500	0.942

*Note: When B_1/B_2 is constant and α level constant, then the power will be the same, because $\hat{\lambda}$ values will be generated by the same numbers.

Reliability growth can be planned by a reliability improvement program. For equipments exhibiting infantile failure the statistical test would be change in shape parameter from $0 < B_1 \leq 1.0$ to $B_2 > 1.0$. For fatigue and cyclic failures one can improve the equipment by increasing the scale parameter from θ_1 to θ_2 where $\theta_1 < \theta_2$ assuming the shape parameter is constant $B_1 = B_2$. Another approach is a change in shape parameter B_1 to B_2 where $B_1 > B_2$ and B_1 in fatigue region.

<u>Improvement Decisions</u>

Condition	Change	Shape	Scale
Infantile Failure	Increase B	$0 < B_1 < 1$ to $B_2 > 1$	θ_1, θ_2
Cyclic Failure	Increase θ (MTBF)	$B_1 = B_2$	$\theta_1 < \theta_2$
Fatigue	Decrease B Increase θ	$B_1 > B_2$	$\theta_1 < \theta_2$
	Increase θ	$B_1 = B_2$	$\theta_1 < \theta_2$

The risks involved with the decisions table are two fold, (1) risk of type I error is α and (2) risk of a type II error is 1.0 - power.

Since θ_1, θ_2, B_1, B_2 are unknown parameters the actual value is determined within intervals by confidence interval estimates. But changes from θ_1 to θ_2 may be tested with the McCool test and changes from B_1 to B_2 may be tested with the maximum likelihood ratio test of the last section. The power the maximum likelihood ratio test seems to be high in the

region where $B_1 \neq B_2$ is appropriate for maintenance decisions, that is $B_1/B_2 \geq 2$, B_1 from 0.3 to 5.0, and fairly small sample sizes.

Appendix

Reliability Growth by K Steps

The two sample change of shape parameter was discussed in the text above. However, the improvement may involve more effort than one reliability fix and test cycle. Suppose a set of K life tests, each obtaining $n_i (i = 1,\ldots,k)$ sample observations. Then the null hypothesis to be tested is

$$H_0: B_1 = B_2 = \ldots = B_k \tag{A1}$$

and the alternative hypothesis H_1 is that H_0 is false. For this test the likelihood ratio test can be derived in a similar way:

$$\hat{\lambda} = \prod_{j=1}^{k} (\hat{\theta}_j/\hat{\hat{\theta}}_j)^{n_j} (\hat{\hat{B}}_1/\hat{B}_j)^{n_j} \prod_{i=1}^{n_j} x_{ij}^{\hat{\hat{B}}_1 - \hat{B}_j} \tag{A2}$$

where $\hat{\hat{B}}_1$, $\hat{\hat{\theta}}_j$ $(j = 1,\ldots,k)$ are M.L. estimators under H_0 and $\hat{B}_j$, $\hat{\theta}_j$ are maximum likelihood estimates for each sample. To obtain the MLE estimates under H_1

$$\hat{\theta}_j,\ \hat{B}_j \qquad j = 1,\ldots,k$$

the equations to follow must be solved

$$n_j/\hat{B}_j + \sum_{i=1}^{n_j} \ln x_{ij} - \sum_{i=1}^{n_j} x_{ij}^{\hat{B}_j} \cdot \frac{\ln x_{ij}}{\hat{\theta}_j} = 0 \tag{A3}$$

$$\hat{\theta}_j = \sum_{i=1}^{n_j} x_{ij}^{\hat{B}_j}/n_j \text{ for } j = 1,2,\ldots,k \tag{A4}$$

and to obtain the MLE estimates under H_0

$$\sum_{j=1}^{k} n_j/\hat{\hat{B}}_1 + \sum_{j=1}^{k} \ln \prod_{i=1}^{n_j} x_{ij} - \sum_{j=1}^{k} \sum_{i=1}^{n_j} x_{ij}^{\hat{\hat{B}}_1} \cdot \frac{\ln x_{ij}}{\hat{\hat{\theta}}_j} = 0 \tag{A5}$$

where

$$\hat{\hat{\theta}}_j = \sum_{j=1}^{n_j} x_{ij}^{\hat{\hat{B}}_1}/n_j \text{ for } j = 1,2,\ldots,k. \tag{A6}$$

It is clear from (A1) and the argument discussed in the text that the critical points are independent of B_1 and depend only on n_1, n_2, ..., n_k and α.

Also the Power of the test is independent of the θ_1, θ_2, ..., θ_k parameters.

<u>References</u>

[1] Bilikam, John E., "F111 Reliability, Maintainability and Availability as Determined from SEDS Data", OPS Analysis, ASD/XRO, February 1971, WPAFB OH 45433.

[2] Bilikam, J. E., "Conditions for Burn-In", ASD/ENYZ, March 1972, WPAFB OH 45433.

[3] Bilikam, Edward and Moore, Albert H., "Estimation of an Aeronautical Equipment Mission Reliability", presented at the Mini-Symposium on New Avenues for Progress in Aerospace Research Technology and Systems, AIAA and Wright Aeronautical Laboratory, March 1976. Submitted IEEE Transactions on Reliability.

[4] Harter, H. L. and Moore, A. H., "Maximum Likelihood Estimation of the Parameters of Gamma and Weibull Populations from Complete and Censored Samples", Technometrics, Vol. 7, pp. 639-643, November 1965.

[5] McCool, John I., "Inferential Techniques for Weibull Populations II", Aerospace Research Laboratories, AFSC, June 15, 1975, ARL-TR-75-0233.

[6] Petrick, George S., "Discriminating Between Weibull Distributions By Means of a Likelihood Ratio Test", Air Force Institute of Technology, Thesis, GOR/MA/75-2 (December 1975), WPAFB OH 45433.

[7] Wingo, D. R. "Maximum Likelihood Estimation of Parameters of the Weibull Distribution by Modified Quasilinearization", IEEE Transactions on Reliability, R-21, pp. 89-93 (1972).

Statistical Analysis and Evaluation of Data Base for Simulation Studies

by

Andrew W. Lai

College of Business And Administration

Wright State University

Dayton, Ohio 45431

STATISTICAL ANALYSIS AND EVALUATION OF DATA BASE FOR SIMULATION STUDIES

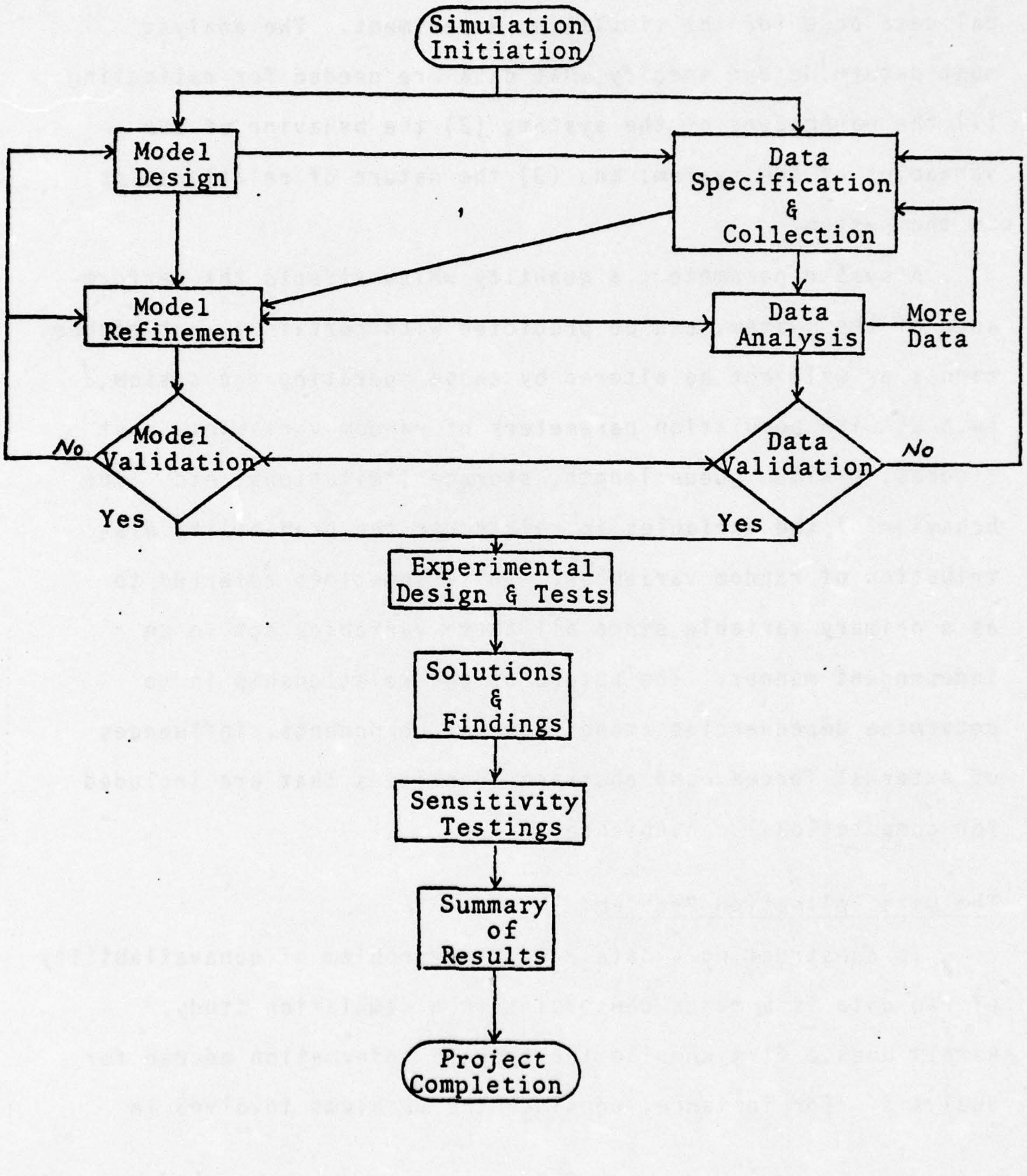
Much of the effort expended in the field of simulation has been directed toward problems of model building, computational methods, and general simulation languages. Little research has been done on problems of data base construction for simulation studies. As shown in Figure 1, data base construction is an indispensable part of a successful simulation study. To a large degree model building depends on the data analysis in terms of relevancy and accuracy of assumptions of the designated model. Although the level and type of data base requirements may vary for different simulation models, the need for a sound data base is the same. This paper will first discuss briefly the specification of data requirements, then explore the data collection problem, and finally suggest the statistical techniques for data analysis and data validations.

Specification of Data Requirements

The data requirements for simulation studies can be grouped into two major categories: (1) historical data for estimating the values of parameters, the behavior of variables, and the nature of relationships; and (2) a data recording system that will capture data at their source and keep all pertinent data records updated. The data recording system is an important part of model building, particularly in simulators

Figure 1

SEQUENCE OF ACTIVITIES FOR A SIMULATION STUDY



designed for operational control of operating systems. It will not be dealt with here in this paper.

The degree of abstraction of the mathematical model representing the system under study will determine the historical data base for the simulation experiment. The analyst must determine and specify what data are needed for estimating: (1) the parameters of the system; (2) the behavior of the variables of the system; and (3) the nature of relationships in the system.

A system parameter, a quantity which affects the performance of the system, can be predicted with certainty, but either cannot or will not be altered by those operating the system, such as with population parameters of random variables, cost figures, maximum queue length, storage limitations, etc. The behavior of the variables is related to the probability distribution of random variables. It is sometimes referred to as a primary variable since all these variables act in an independent manner. The nature of the relationship is to determine dependencies among internal components, influences of external forces, and abstract identities that are included for computational convenience.

The Data Collection Problem

In constructing a data base, the problem of nonavailability of raw data is a major constraint in a simulation study. Rarely does a firm compile the type of information needed for analysis. For instance, consider the problems involved in

attempting to estimate the cost of a lost sale or the cost of a backorder. Here one must estimate the cost of intangibles, such as lost goodwill which may be converted into the loss of business in the future. Thus, the problem of nonavailability of the data is a major factor in constructing a viable data base.

The second related problem is that the data being collected is seldom in a form which is suitable for analysis. For example, in undertaking a freight study, the analyst needs accurate information on the rate formulations in terms of product, quantity, size, distance, modes of transportation and so forth. Unfortunately, the information usually available to him is the rate structure which is so complicated that it is difficult, if not impossible, to analyze.

Another problem which impairs construction of the data base is that the data may be obtained from various sources and may be inaccurate and/or imcomplete. An analyst must constantly search for more current data as well as for ways to estimate from, and validate, available data.

The analyst is initially concerned with where he can begin to gather essential data, regardless of the type of study he is conducting. In studies of business operating systems, the company's sales invoices or bills of lading provide an excellent starting point to begin gathering the data. A list of Internal Sources of Data is presented in Table 1, and External Sources of Data in Table 2.

Table 1

INTERNAL SOURCES OF DATA

Source	Type of Data
Accounting Records	Overhead charges, taxes, utility costs, transportation costs, profits, profit margins.
Engineering Records	Material specfications, equipment performances, technological ordering.
Inventory Records	Raw materials, in-process, finished goods.
Maintenance Records	Reliability of components of system; machine downtime, failure rate, repair-time and waiting-time distributions; frequencies of external perturbations.
Personnel Records	Salaries, absenteeism, medical records, skill classifications, labor availability, labor turnover.
Production Control Records	Schedule status, setup times, routing and sequencing information, assembly-line balancing, loading, decision rules.
Quality Control Records	Machine performance, scrap, defectives, probability distributions of attributes and measurements, effect of machine age.
Tool Crib Records	Frequency of demands, waiting-time distributions, frequency of tool breakage, lifetime distributions of tools.
Warehouse Records	Frequency of issuing materials and supplies, peak periods of activity, waiting-time distributions, frequency of stockouts, individual and aggregate rate of expenditure of materials and supplies.

Table 2

EXTERNAL SOURCES OF DATA

Customers
Governmental agencies
Industrial and trade organizations
Insurance Companies
Management consultants
Market research firms
Publications in journals of professional organizations
Standard data
Universities
Vendors and suppliers

Statistical Techniques in Data Base Analysis

1. Sampling Considerations

The development of the sample design is a critical stage in the establishment of a data base. The two most important features of a sample are its size and the manner in which it was selected. Much of the study of sampling statistics concerns the determination of these two characteristics. As expected, this determination is based upon the specific conditions prescribing the purpose of the sample. For instance, the sample in the business system study has to be large enough to be representative of a product group's distribution pattern in each geographical subdivision, yet small enough to be relatively inexpensive to collect and prepare by the analyst.

A simple random sampling may be simple to use but one must carefully evaluate the result. While the population as a whole may be adequately represented, the sample may not be representative of each product group and geographical subdivision due to the chance variation that some subdivisions may have an insufficient number of observations to provide adequate information.

When many subdivisions are involved stratified random sampling should be selected. The advantages of this procedure follow from the fact that the population is divided into subdivisions. Thus, only a relatively small number of observations are needed to determine the characteristics of each subdivision.

2. Sales Forecasting

After the sample has been designed and taken, the analyst can proceed to sales forecasting. There are two levels of

sales forecasting needed in business system studies: One level of sales forecasting is the aggregate sales forecast, such as monthly, quarterly, or even yearly sales forecasts for the product by market regions. The second level of forecasting is detail forecasting, which is critical for a simulation study, such as frequency of order arrival and order size for each product by market region. In order to project these two level of sales forecasting trend and seasonality are necessary.

a. Trend of the Sales. Most companies have an extensive amount of information on trends established for their products with which the analyst can predict a sales forecast. If the trend data are not available or are insufficient, one can then obtain the trend prediction by analyzing the historical sales data. The simplest method to determine the trend is that of least squares.

The use of a least squares method does not eliminate all subjectivity from the analysis of time series. Among other things, a decision must be made on the number of years to use, what adjustments are to be made, and whether to fit a straight line in preference to other types of curves. However, the location of a line for a given series of data is determined because only one line satisfies the above least squares criterion.

Another way to determine the trend of previous sales is the method of moving averages. Moving averages serves to

eliminate some of the variability not attributable to the trend in an attempt at its identification. (See Figure 3). One method of identifying the appropriate trend curve is to look at certain characteristics of a curve's moving averages. The reader is referred to the book, <u>Mathematical Trend Curves</u> by Gregg, Hossell, and Richardson for further discussion and examples of the process of fitting trend curves to time series data.

A third method of determining trend lines is that of exponential smoothing. The chief advantage of exponential smoothing is that only a minimal amount of storage in the computer is required; only five computer instructions are necessary and can be used repeatedly. For each degree of smoothing, the data to be stored are quite brief; the computations are quick and simple, and the smoothing constant can be adjusted at will as current information indicates a need for change. This method has been widely applied in simulation studies.

Once the trend line has been determined, it is necessary to make inferences of future demand. Having obtained the quation of a trend, it can be used to extrapolate by means of estimating a value which lies beyond the range of values on the basis of which an equation was originally obtained, and thus forecast future sales.

Extrapolating from trends is a necessary, though highly speculative, procedure whose success depends on many factors. The basic question is whether the forces which have operated

Figure 2

Linear Trend Fitted by the Least Squares Method

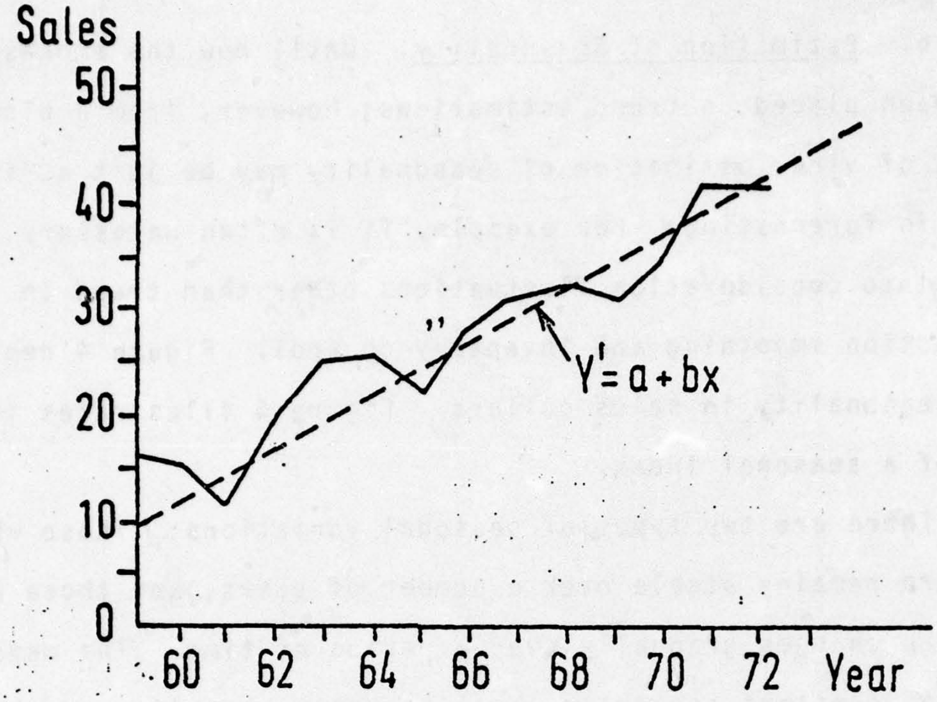
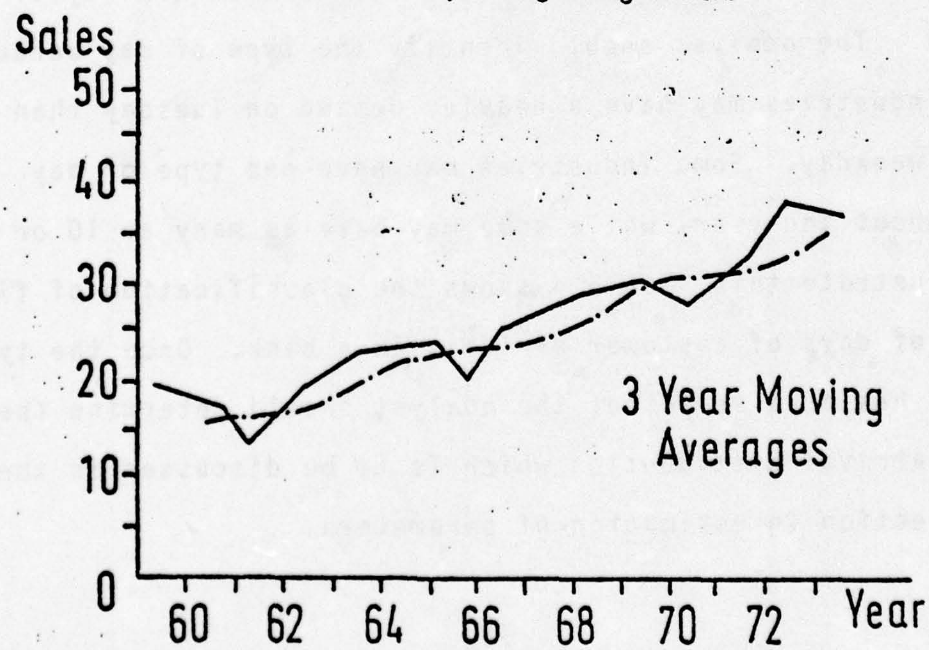


Figure 3

Trend Fitted by Moving Average Method



in the past will continue to operate in precisely the same manner in the future.

b. Estimation of Seasonality. Until now the emphasis has been placed on trend estimations; however, from a planning point of view, estimation of seasonality may be just as important in forecasting. For example, it is often necessary to take into consideration fluctuations other than trend in production smoothing and inventory control. Figure 4 depicts the seasonality in sales dollars. Figure 5 illustrates the use of a seasonal index.

There are two types of seasonal variations: Those whose pattern remains stable over a number of years, and those whose pattern changes gradually over a period of time. The measurement of constant seasonals involves rather complicated theoretical and practical considerations.

c. Types of Day. There are many factors which affect the number of orders received each day--for example, type of day, seasonality, trend, as well as the random quality of the orders. The analyst should identify the type of day because some industries may have a heavier demand on Tuesday than any other weekday. Some industries may have one type of day throughout the year, while some may have as many as 10 or more. To illustrate this, Table 4 shows the classification of five types of days of customer arrivals in a bank. Once the type of day has been selected, the analyst should determine the order arrival distrubution which is to be discussed in the next section in estimation of parameters.

Figure 4

Sales in Dollars by Month

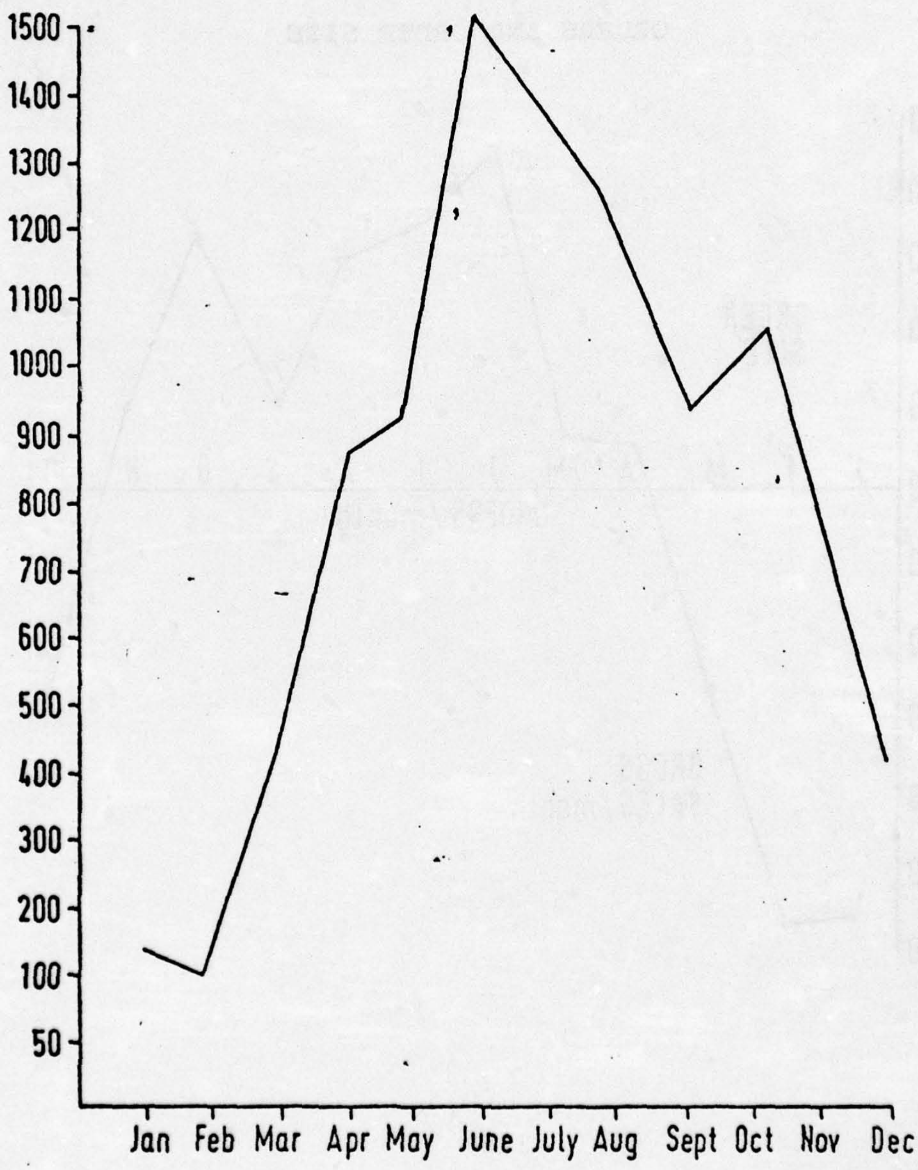


Figure 5

MONTHLY INDICES FOR GROSS SALES, NUMBER
,
ORDERS AND ORDER SIZE

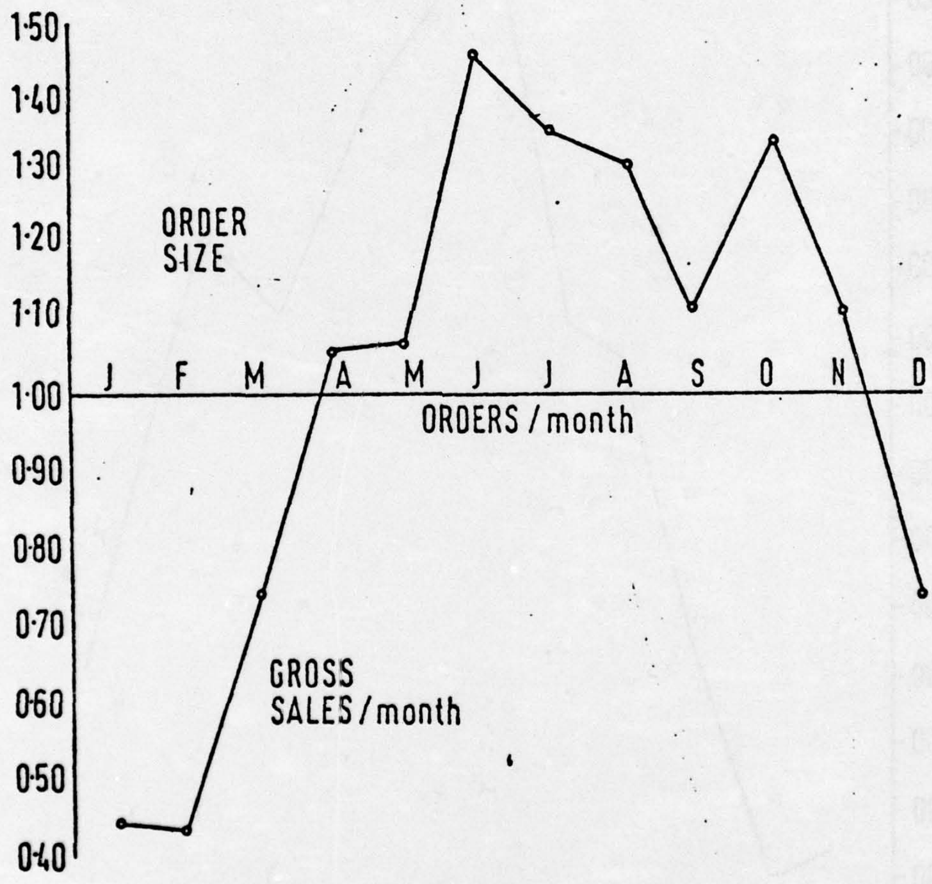


Table 3

Types of Day of Customers Arrival

Type 1	The third Friday of the month
Type 2	All other Fridays
Type 3	The first day of the month if it is a Tuesday, Wednesday or Thursday
Type 4	All Mondays and the first two Tuesdays and Thursdays of the month
Type 5	Wednesdays, and the last two or three Tuesdays and Thursdays of the month

3. Estimation of Parameters

The discussion which follows attempts to point out some useful methods for estimating system parameters. This treatment by no means exhausts all available methods for parameter estimation. The reader should consult a text on mathematical statistics or statistical inference for a more sophisticated treatement.

Three steps are involved in defining the distribution of a random variable. First, one must estimate the general form of the distribution. The second step is to estimate the parameters of the hypothesized distribution. The last step is to determine whether or not the hypothesized distribution adequately represents the random variable in question. If one concludes that the hypothesized distribution in unreliable, the analyst should repeat this procedure, starting with a search for a new distribution.

a. <u>Identifying the Distribution.</u> With very few exceptions one cannot make a reasonable guess regarding the distribution of a random variable until data has been collected which can be used as a guide. Plotting the relative frequency distribution of the random variable under study is often helpful in determining its probability mass function or density function. For instance, from an order record of 203 days one can construct a frequency and probaiblity distribution of the orders arrival such as in Table 4. Figure 6 is transformed from the Table 4 probability distribution into a histogram distribution, which may suggest that the distribution follows that of a Poisson distribution. Also shown in Figure 7, order size may suggest a negative exponential distribution.

b. <u>Estimation of Distribution Parameters</u>. Once the analyst has identified one or more distribution classes which he feels adequately represents the variable he is studying, he then determines the numerical values of the distribution parameters in order to reduce the distribution class to a specific distribution. When the hypothesized distribution is a function of two parameters, he can usually estimate these parameters from the sample mean and sample variance. Table 5 presents common statistical distributions and their repsective parameters, and general form which may aide the analyst in determining the probability distribution and estimating the associated parameters.

c. <u>Validating and Testing the Probability Distribution</u>. Having hypothesized that a random variable is characterized by a specific probability distribution, one is left with the task

Table 4

Probability Distribution of Orders Arrival

Orders Arrival	Frequency	Probability	Cumulative Probability
14	19	0·094	0·094
15	42	0·207	0·301
16	44	0·217	0·518
17	41	0·202	0·720
18	26	0·128	0·848
19	11	0·056	0·904
20	9	0·042	0·946
21	4	0·020	0·966
22	3	0·015	0·981
23	2	0·010	0·991
24	1	0·005	0·996
25	1	0·004	1·000
Total	203	1·000	

Figure 6

Relative Frequency of Orders Arrival

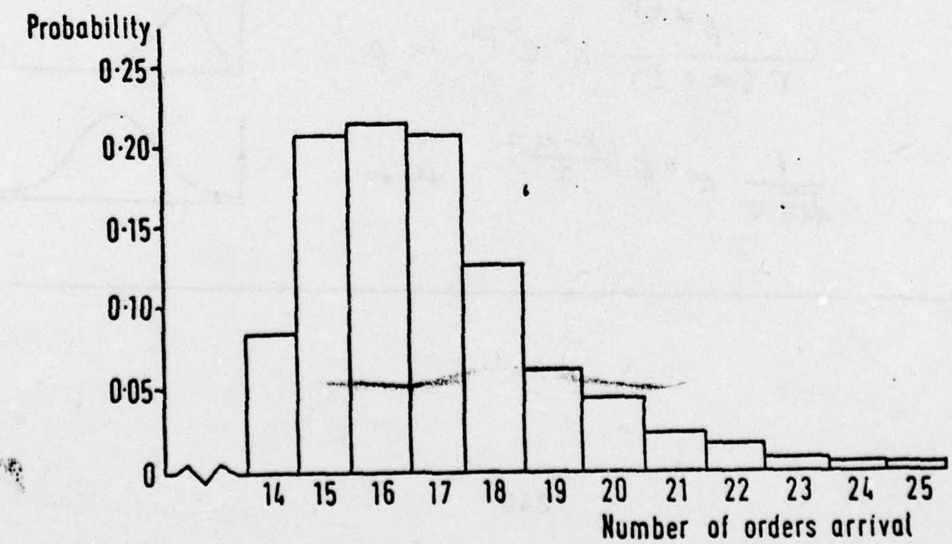


Table 5

Table of Common Statistical Distribution

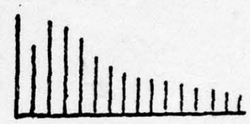
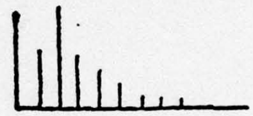
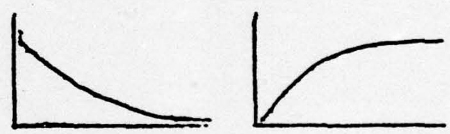
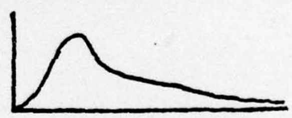
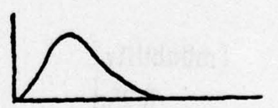
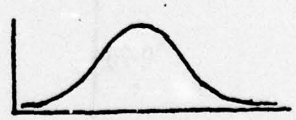
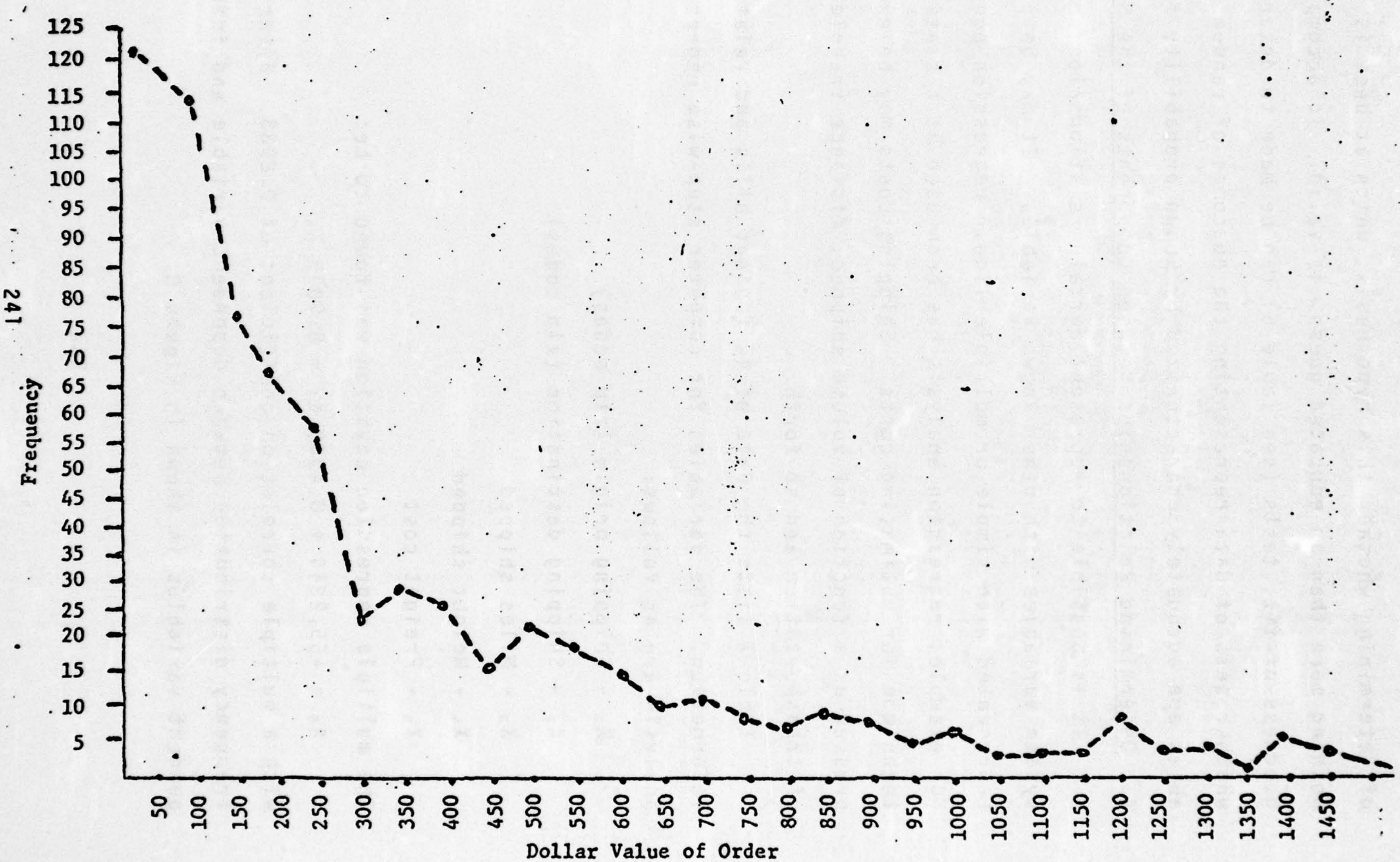
Distribution	Density Function	Parameters	General Form
Binomial	$\binom{n}{r} p^r (1 - p)^{n-r}$	n,p	
Poisson	$\frac{\lambda^r}{r!} e^{-\lambda}$	λ	
Exponential	$\frac{1}{\beta} e^{-(\frac{t}{\beta})}$	β	
Gamma	$\frac{(\frac{t}{\mu})^{r-1} e^{-(\frac{t}{\mu})}}{(r - 1)!}$	$r, \frac{1}{\mu}$ (α, β)	
Beta	$\frac{\beta^{\alpha+1}}{\Gamma(\alpha + 1)} x^{\alpha} e^{-\beta x}$	α, β	
Normal	$\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$	μ, σ	

Figure 7

SAMPLE DISTRIBUTION OF ORDERS BY DOLLAR SIZE



of determining whether this hypothesis, which at best is usually nothing more than an educated guess, is valid. To accomplish goodness-of-fit, tests (see Table 6) can be made to determine whether sets of data representing the outcomes of random variables are adequately characterized by known probability functions.

4. Determining Relationships Between Components of the System

It is possible to represent certain relationships between system variables with other known variables. It may be properly represented with simple or multiple linear regression equations. For example, regression analysis has been used as a statistical technique for estimating costs. Shipping costs may be expressed as a function of volume shipped, distance traveled, mode of transportation and so forth.

Table 7 lists the data of 44 freight bills and related information. The variables for computer step-wise regression analysis are as follows:

X_1 - Shipping origin (zip codes)

X_2 - Shipping destination (zip codes)

X_3 - Miles shipped

X_4 - Weight shipped

X_5 - Freight cost

The multiple regression equation was found to be:

$$X_5 = -55.2844 + 0.4943\, X_3 + 0.0079\, X_4$$

with a multiple correlation coefficient of 0.8983. Bivariate frequency distribution between dependent variable and independent variables is shown in Figure 8.

Table 6

COMMONLY USED STATISTICAL TECHNIQUES FOR TEST OF GOODNESS-OF-FIT

Techniques	Applications
Analysis of Variance	Test the hypothesis that the mean (or variance) is equal to the mean (or variance) of the corresponding observed series
Chi-Square Test	Test the hypothesis that the set of generated data has the same frequency distribution as a set of observed historical data
Kolmogorov-Smirnov Test	This test may be used as an alternative to a chi-square test. It has two advantages over chi-square: 1. It is more powerful than chi-square test. 2. It can be applied to a very small sample.
Regression Analysis	Regress actual series on the generated series and test the resulting regression equations
Spectral Analysis	The analysis provides a means of objectively comparing time series generated by a computer model with observed time series
Nonparametric Tests	There are many other nonparametric tests besides kolmogorov-Smirnov test are available for test of goodness-of-fit. For reference, see Siegel's Nonparametric Statistics.

Table 7

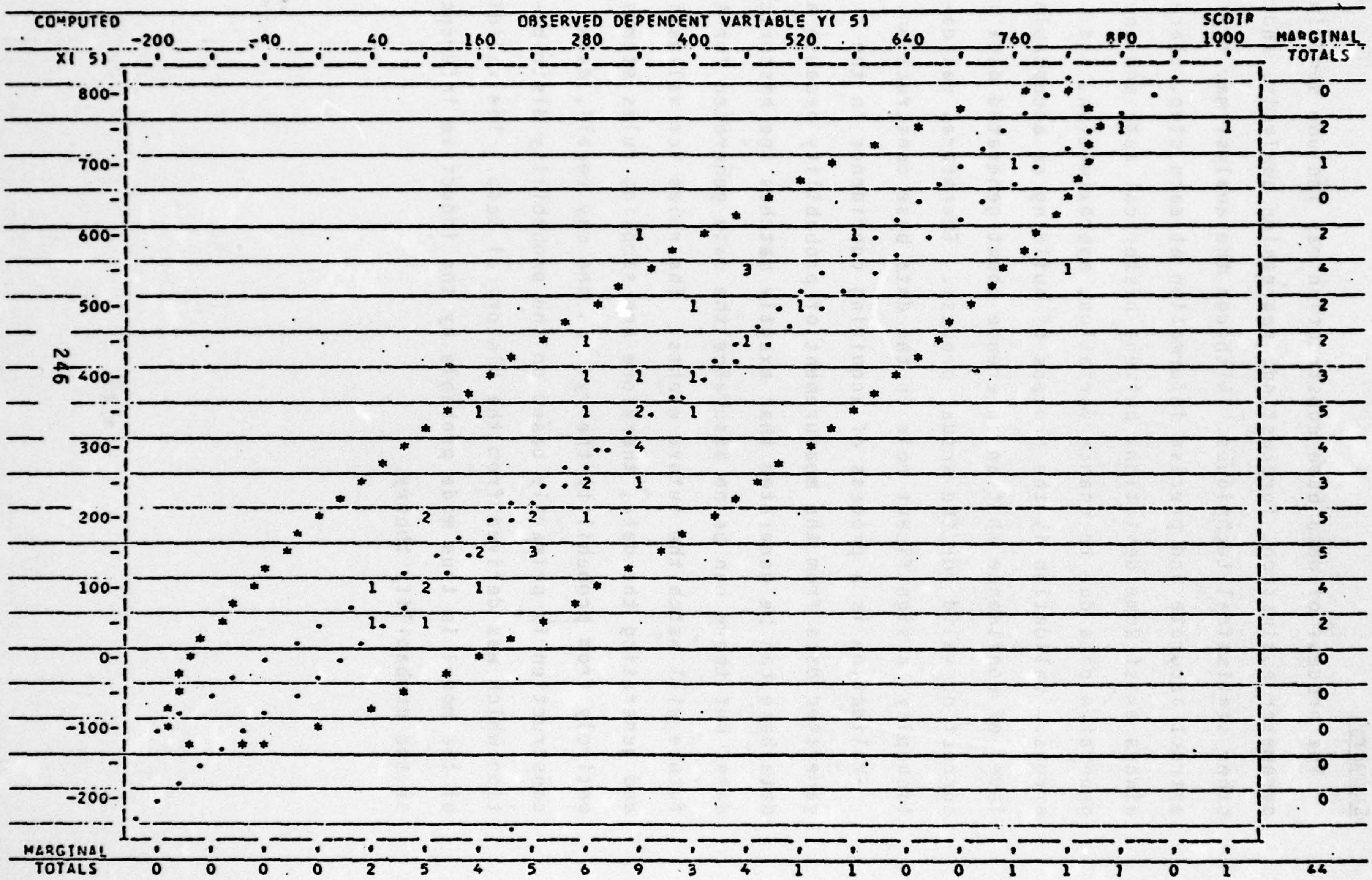
Data of a Sample of 44 Freight Bills

OSU/CCONI STEP-WISE REGRESSION PROGRAM IV MAIN PROBLEM 2 TL 55 REGRESSION PROBLE

DATA MATRIX		X_1	X_2	X_3	X_4	X_5	
ORIGINAL	1	44100.0	30000.0	707.0000	10368.0	272.5000	55.0000
ORIGINAL	2	44100.0	44000.0	21.0000	10250.0	78.2200	55.0000
ORIGINAL	3	44100.0	14200.0	185.0000	10319.0	167.0100	55.0000
ORIGINAL	4	44100.0	19500.0	364.0000	10722.0	233.7400	55.0000
ORIGINAL	5	44100.0	12000.0	330.0000	10759.0	212.0100	55.0000
ORIGINAL	6	44100.0	45900.0	151.0000	10947.0	114.2800	55.0000
ORIGINAL	7	44100.0	2900.0000	471.0000	10997.0	246.9200	55.0000
ORIGINAL	8	44100.0	24000.0	470.0000	11073.0	239.4500	55.0000
ORIGINAL	9	44100.0	14000.0	220.0000	12722.0	201.0100	55.0000
ORIGINAL	10	60400.0	60400.0	20.0000	12928.0	51.5500	55.0000
ORIGINAL	11	44100.0	30000.0	708.0000	13456.0	267.1500	55.0000
ORIGINAL	12	44100.0	21200.0	344.0000	13947.0	319.3899	55.0000
ORIGINAL	13	44100.0	44800.0	52.0000	14307.0	124.4700	55.0000
ORIGINAL	14	44100.0	43100.0	170.0000	14900.0	205.7200	55.0000
ORIGINAL	15	44100.0	43100.0	174.0000	14900.0	205.7200	55.0000
ORIGINAL	16	44100.0	13100.0	320.0000	15157.0	268.8000	55.0000
ORIGINAL	17	44100.0	44000.0	24.0000	15134.0	70.0000	55.0000
ORIGINAL	18	44100.0	15000.0	115.0000	17450.0	190.0000	55.0000
ORIGINAL	19	44100.0	30300.0	709.0000	17900.0	441.6001	55.0000
ORIGINAL	20	44100.0	11700.0	500.0000	17159.0	354.8000	55.0000
ORIGINAL	21	44100.0	19700.0	391.0000	21120.0	331.2000	55.0000
ORIGINAL	22	44100.0	1500.0000	593.0000	21224.0	291.2000	55.0000
ORIGINAL	23	44100.0	11700.0	491.0000	21553.0	350.7999	55.0000
ORIGINAL	24	44100.0	20000.0	344.0000	21572.0	352.7600	55.0000
ORIGINAL	25	44100.0	55415.0	737.0000	21927.0	371.2400	55.0000
ORIGINAL	26	44100.0	77000.0	1290.0000	21930.0	955.9999	55.0000
ORIGINAL	27	44100.0	48100.0	167.0000	24299.0	97.2000	55.0000
ORIGINAL	28	44100.0	53500.0	459.0000	24301.0	298.4501	55.0000
ORIGINAL	29	44100.0	0.0	287.0000	25440.0	313.2600	55.0000
ORIGINAL	30	60600.0	60000.0	20.0000	26390.0	158.2900	55.0000
ORIGINAL	31	44100.0	44400.0	50.0000	26494.0	101.2500	55.0000
ORIGINAL	32	44100.0	30100.0	698.0000	29014.0	499.0100	55.0000
ORIGINAL	33	44100.0	55400.0	737.0000	29057.0	490.0000	55.0000
ORIGINAL	34	44100.0	72100.0	125.0000	31202.0	405.6197	55.0000
ORIGINAL	35	44100.0	2100.0000	671.0000	34560.0	493.0100	55.0000
ORIGINAL	36	44100.0	2100.0000	671.0000	39215.0	445.9700	55.0000
ORIGINAL	37	44100.0	55400.0	747.0000	39274.0	544.2799	55.0000
ORIGINAL	38	44100.0	45800.0	151.0000	39175.0	316.4399	55.0000
ORIGINAL	39	44100.0	49200.0	149.0000	39292.0	140.0000	55.0000
ORIGINAL	40	44100.0	60400.0	235.0000	45442.0	276.4400	55.0000
ORIGINAL	41	44100.0	2100.0000	651.0000	54440.0	754.2101	55.0000
ORIGINAL	42	44100.0	1700.0000	601.0000	61535.0	901.1499	55.0000
ORIGINAL	43	44100.0	19000.0	222.0000	64725.0	941.4299	55.0000
ORIGINAL	44	44100.0	45400.0	202.0000	69244.0	356.0601	55.0000

Figure 8

BIVARIATE FREQUENCY DISTRIBUTION BETWEEN COMPUTED AND OBSERVED DEPENDENT VARIABLE Y(5) OF PROBLEM NUMBER 2.02
FIT OF ONLY TWO VARIABLES, WEIGHT AND DIS



Summary

The process of data base construction may include sampling, parameter estimation, forecasting, regression analysis, and other statistical techniques. Although the analyst may try to extract accurate and precise information at each step, there always exist some deviations betwene historical data and the generated data due to chance variation, approximation, and errors. Validation is the process of building an acceptable level of confidence that an inference about generated data is correct or valid for the actual process. Therefore, validation plays a significant role in the data base construction.

Validation is a process of acquiring confidence in the generated data from the measurement of probability because a data base can be generated that exactly matches the historical data, but there can be no assurance the data generated for the future will match the future events. The rules for validating and generating the data, therefore are sampling rules stemming entirely from probability theory. As one may recall, data construction is primarily based on the probability distribution which was derived from the historical data. The validity of the model is thus made probable by the inductive inference of the probability theory.

BIBLIOGRAPHY

1. Benston, George, "Multiple Regression Analysis of Cost Behaviour", Accounting Review, October, 1966, pp. 657-72.

2. Bonini, C. P., R. K. Jaedicke, and H. M. Wagner, eds., Management Controls. McGraw-Hill Book Company, New York, 1964.

3. Brown, Robert Goodell, Smoothing, Forecasting and Prediction, Englewood Cliffs, N. J., Prentice-Hall, 1963.

4. Comiskey, Eugene, "Cost Control by Regression Analysis", Accounting Review, April, 1966, pp. 235-8.

5. Conway, R. W. "Some Tactical Problems in Digital Simulation", Management Science, 10, October, 1963, 47.

6. Croxton, Frederick E., and Dudley, J. Cowden, Applied General Statistics, 2nd ed., Englewood Cliffs, N. J.: Prentice-Hall, 1955.

7. Deming, W. Edwards, Sample Design in Business Research, New York: Wiley, 1960.

8. Hicks, C. R., Fundamental Concepts in the Design of Experiments. Holt, Rinehart & Winston, Inc., New York, 1964.

9. Lai, Andrew W., and LaLonde, Bernard J., An Analysis and Evaluation of Data Base Requirements for Physical Distribution Studies, International Journal of Physical Distribution, Bradford, Yorkshire, U. K., 1974.

10. Mirham, G. Arthur, "Some Practical Aspects of the Verification and Validation of Simulation Models", Operations Research Quarterly, March, 1972, p. 21.

11. Mize, J. H., and Cox, J. Grady, Essentials of Simulation, Englewood Cliffs, N. J.: Prentice-Hall, 1968.

12. Naylor, T. H., Balintfy, J. L., Burdick, D. S., and Chu, k., Computer Simulation Techniques. John Wiley & Sons, Inc., New York, 1966.

13. Siegel, Sidney, Nonparametric Statistics, McGraw-Hill, New York, 1956.

14. Schmidt, J. W., and R. E. Taylor, Simulation and Analysis of Industrial Systems, Richard D. Irwin, Inc., Homewood, Illinois, 1970.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFIT TR 76-15	2. GOVT ACCESSION NO. / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): PROCEEDINGS OF THE SYMPOSIUM ON APPLICATION OF DECISION THEORY TO PROBLEMS OF DIAGNOSIS AND REPAIR	5. TYPE OF REPORT & PERIOD COVERED / 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Norman Keith Womer, Editor	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: AFIT/EN, Wright-Patterson AFB, OH 45433	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: AFIT/EN, Wright-Patterson AFB, OH 45433	12. REPORT DATE: October 1976 / 13. NUMBER OF PAGES: 254
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED / 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Approved for public release; IAW AFR 190-17

JERRAL F. GUESS, Capt, USAF
Director of Information

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Decision Theory
Maintenance
Repair
Statistics

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The papers presented at this symposium focus on the analysis of decision making in an environment where the diagnosis and repair of complex mechanisms requires the use of still imperfect measurements of several factors. These analyses take several forms, among these are: (1) estimating the effects of incorrect decisions on costs, reliability and other elements of the loss function, (2) determining the variability of tests and test procedures, (3) determining the best sequence of tests, and (4) simulating the effects of

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

alternative decision criteria on equipment repair processes. None of the papers offers a complete approach to decision making in this complex environment. However, each presents an approach to a significant portion of the problem. Our hope is that in discussing and comparing the approaches presented here, our understanding of the problem will be increased.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)